PRINCIPLES OF MATHEMATICAL PHYSICS

HOUSTON

SECOND EDITION

53.01
HOU

McGRAW-HILL
BOOK COMPANY

INTERNATIONAL SERIES IN PURE AND APPLIED PHYSICS

LEONARD I. SCHIFF, CONSULTING EDITOR

ADVISORY EDITORIAL COMMITTEE: E. U. Condon, George R. Harrison
Elmer Hutchisson, K. K. Darrow

PRINCIPLES OF MATHEMATICAL PHYSICS

INTERNATIONAL SERIES IN PURE AND APPLIED PHYSICS

LEONARD I. SCHIFF, *Consulting Editor*

Allis and Herlin Thermodynamics and Statistical Mechanics
Becker Introduction to Theoretical Mechanics
Clark Applied X-rays
Edwards Analytic and Vector Mechanics
Evans The Atomic Nucleus
Finkelnburg Atomic Physics
Ginzton Microwave Measurements
Green Nuclear Physics
Gurney Introduction to Statistical Mechanics
Hall Introduction to Electron Microscopy
Hardy and Perrin The Principles of Optics
Harnwell Electricity and Electromagnetism
Harnwell and Livingood Experimental Atomic Physics
Harnwell and Stephens Atomic Physics
Houston Principles of Mathematical Physics
Houston Principles of Quantum Mechanics
Hughes and DuBridge Photoelectric Phenomena
Hund High-frequency Measurements
Kemble The Fundamental Principles of Quantum Mechanics
Kennard Kinetic Theory of Gases
Marshak Meson Physics
Morse Vibration and Sound
Morse and Feshbach Methods of Theoretical Physics
Muskat Physical Principles of Oil Production
Read Dislocations in Crystals
Richtmyer, Kennard, and Lauritsen Introduction to Modern Physics
Schiff Quantum Mechanics
Seitz The Modern Theory of Solids
Slater Introduction to Chemical Physics
Slater Microwave Transmission
Slater Quantum Theory of Matter
Slater and Frank Electromagnetism
Slater and Frank Introduction to Theoretical Physics
Slater and Frank Mechanics
Smythe Static and Dynamic Electricity
Squire Low Temperature Physics
Stratton Electromagnetic Theory
Thorndike Mesons: A Summary of Experimental Facts
Townes and Schawlow Microwave Spectroscopy
White Introduction to Atomic Spectra

The late F. K. Richtmyer was Consulting Editor of the series from its inception in 1929 to his death in 1939. Lee A. DuBridge was Consulting Editor from 1939 to 1946; and G. P. Harnwell from 1947 to 1954.

PRINCIPLES OF MATHEMATICAL PHYSICS

By WILLIAM V. HOUSTON

PRESIDENT, RICE INSTITUTE OF TECHNOLOGY

SECOND EDITION

NEW YORK TORONTO LONDON

McGRAW-HILL BOOK COMPANY, INC.

1948

PRINCIPLES OF MATHEMATICAL PHYSICS

Copyright, 1934, 1948, by the McGraw-Hill Book Company, Inc. Printed in the United States of America. All rights reserved. This book, or parts thereof, may not be reproduced in any form without permission of the publishers.

x5890027x
HARRIS COLLEGE
PRESTON
530.15 Hou
3251

THE MAPLE PRESS COMPANY, YORK, PA.

PREFACE TO THE SECOND EDITION

The first edition of this book was written as a text for a course designed to give the student some competence in the techniques of classical mathematical physics and some confidence in his ability to read technical papers in that field.

The second edition has been amplified to better serve this end without any very wide deviation from the fundamental method of approach.

Chapters on mathematical methods alternate with those in which the methods are applied to physical problems, and everywhere the emphasis is on the drawing of quantitative conclusions from carefully stated laws.

The amplification has consisted largely of the inclusion of numerous illustrative examples completely or partly developed in the text, the addition of drawings to clarify the text, and some minor rearrangements of material to provide what seems after use a more coherent order of presentation. A few sections dealing with matters not actually essential to the context have been omitted.

The chapters on electricity and magnetism have been considerably expanded and revised in the effort to present a precise formulation of this conceptually difficult subject. Emphasis has been laid on the similarities of, and the differences between, the vectors **D** and **E** and between **B** and **H**. The point of view presented is believed to be especially helpful in understanding the electrical properties of matter.

Some minor changes have been made in the problems presented for solution by the student. Some of these have been worked out and included in the text, and new ones have been added. These continue to be the backbone of the course. Mathematical physics, obviously, is an art whose mastery can be attained only by extensive practice.

For many valuable suggestions concerning the treatment,

the author is indebted to numerous colleagues, as well as to many of the students who have worked through the text. He is especially obligated to Profs. Carl D. Anderson and Wm. A. Fowler, who have given him the benefit of their teaching experience; to Dr. Leverett Davis, of the California Institute of Technology, who worked over much of the manuscript with him; and to Drs. Charles F. Squire and J. R. Risser, of the Rice Institute, who read much of the proof on electricity and magnetism. The wholehearted interest of one's colleagues is one of the pleasant phases of preparing a manuscript of this kind.

WILLIAM V. HOUSTON

HOUSTON, TEX.
July, 1948

PREFACE TO THE FIRST EDITION

This book has been written as the text for a course which I have given for several years to juniors, seniors, and first-year graduate students. The course has been designed to give a working knowledge of the fundamental methods of mathematical physics rather than to give a critical or an exhaustive exposition of the theories of physics. It has been assumed that the students have a thorough knowledge of elementary physics, analytical geometry, and calculus.

The material presented has not been selected according to any rigid plan. Chapters on differential equations and vector analysis have been included since many students are not well prepared in these subjects during their first two years of college work. Mechanics is emphasized because it furnishes the conceptual basis for all physics and because it furnishes good illustrations of nearly all the important mathematical methods. Considerable attention has been given to the study of normal coordinates and normal modes of vibration because of the importance of this type of mathematics in the problems of the quantum theory. A brief treatment of thermodynamics and of statistical mechanics has been included because they each have characteristic methods which are of importance. The treatment of electricity and magnetism provides an introduction to field-theory methods. It has seemed undesirable to include any discussion of quantum mechanics and the methods peculiar to it, since the classical material fully occupies the time available.

In general, the idea has been to emphasize the derivation of results from explicitly stated postulates and to distinguish carefully between such derivation and physical intuition. This plan has not been rigorously followed, however, because it easily becomes too cumbersome in a textbook.

It is commonly accepted as a platitude that a student learns only what he does for himself, but it is not always easy to make

use of this principle. This book has been written as a text, not as a treatise, and as much as possible has been left to the resourcefulness of the student. The physical theory is presented only in the barest outline. The postulates and the definitions are given together with some illustrative examples. The details of the theory are left for the student to fill in by solving the problems. The problems form an integral part of the text, and many important results are given in them. This necessarily means that the book must not merely be read but must be carefully worked through. The text material given, together with the solutions of the problems, should give a fair idea of the subjects treated.

This form of presentation makes severe demands upon both the student and the instructor. It is not to be expected that many students will be able to solve all of the problems, for they require a considerable amount of insight and ingenuity. It can be expected, however, that in making the attempt the student will become familiar enough with the difficulties to understand the solution when he learns it from the instructor or from some other source. The references at the end of each chapter indicate places in which many of the solutions can be found.

It is impossible to acknowledge properly all of the many sources to which I am indebted, but the references at the end of each chapter include a number of those books which have been of importance in establishing my point of view. I am also indebted to Prof. W. R. Smythe and Dr. M. S. Plesset who have read portions of the manuscript, and to Dr. C. B. Crawley who has assisted in the proofreading.

WILLIAM V. HOUSTON

PASADENA, CALIFORNIA
August, 1934

TABLE OF CONTENTS

PREFACE TO THE SECOND EDITION v

PREFACE TO THE FIRST EDITION vii

CHAPTER I. ELEMENTARY DIFFERENTIAL EQUATIONS 1

1. Nature of a Differential Equation and Its Solution 2
2. Solution by the Separation of Variables 5
3. Formulation of the Differential Equation 5
4. Linear Differential Equations 8
5. Definite and Indefinite Integrals 9
6. Simpson's Rule and Numerical Integration 10
7. Differentiation of Integrals 13
8. Second-order Equations That Are Reducible to the First Order 13

CHAPTER II. THE MECHANICS OF PARTICLES 18

1. Newton's Equations of Motion 18
2. The Energy Integral 20
3. Newton's Third Law of Motion and the Momentum Integral 24
4. The Conservation of Angular Momentum 26
5. The Conservation Laws of Mechanics 27
6. The Motion of a Projectile 28
7. Simple-harmonic Motion 30
8. Motion under an Inverse-square Force 32

CHAPTER III. LINEAR EQUATIONS OF ORDER HIGHER THAN THE FIRST 37

1. General Properties of Linear Differential Equations 37
2. Linear Differential Equations with Constant Coefficients 40
3. Solution of the Homogeneous Equation with Constant Coefficients . . . 40
4. Solution of the Nonhomogeneous Equation with Constant Coefficients . 42
5. Complex Numbers 50
6. Complex Functions of Real Variables 53
7. Hyperbolic Functions 54
8. Principles of Superposition and Decomposition 55
9. Equations with Variable Coefficients 58
10. Solutions in Power Series around Ordinary Points 59
11. Solutions in Power Series around Regular Singularities 63

CHAPTER IV. MECHANICS OF VIBRATING PARTICLES 67

1. Damped Vibrations 67
2. Forced Vibrations 72

3. Coupled Oscillators 80
4. Normal Coordinates 83
5. Oscillations under an External Force 85

CHAPTER V. CALCULUS OF VARIATIONS 87

1. The Variation Problem 87
2. Extreme Values of Ordinary Functions 89
3. Extreme Values of an Integral 91
4. The Euler-Lagrange Equation 93
5. Variation Problems with Several Dependent Variables 95
6. Problems with Auxiliary Conditions 96
7. Isoperimetric Problems 98

CHAPTER VI. HAMILTON'S PRINCIPLE 102

1. Derivation of Hamilton's Principle from Newton's Laws 102
2. Illustration of Lagrange's Equation for Conservative Systems 106
3. Problems Involving Constraints 108
4. Problems with Nonconservative Forces 110
5. Hamilton's Canonical Equations of Motion 111
6. The Pendulum 113

CHAPTER VII. THEORY OF VIBRATING SYSTEMS 120

1. General Theory of Normal Coordinates 120
2. Vibrations of a Loaded String 121
3. Normal Coordinates of a Loaded String 126
4. Forced Vibrations of Loaded String 128
5. Approximation to a Continuous String 128
6. Normal Vibrations of a Continuous String with Fixed Ends 129
7. General Solution and Evaluation of Constants 132
8. The Energy of a Vibrating String 133
9. Forced Vibration of a Continuous String 134
10. Expansion of Functions as Series of Orthogonal Functions 136
11. Vibration of a Nonuniform String 138
12. The Variation Problem for Normal Vibrations 141
13. Traveling Waves in a String 142

CHAPTER VIII. VECTOR ANALYSIS 145

1. The Definition of a Vector 145
2. Addition of Vectors 146
3. Orthogonal Components of Vectors 147
4. Multiplication of Vectors 148
5. Differentiation of a Vector with Respect to a Scalar 151
6. Transformation Properties of Vectors 152
7. Linear Vector Functions 154
8. Tensors 155

CHAPTER IX. DYNAMICS OF RIGID BODIES 157

1. Center of Mass and Linear Momentum 157
2. Angular Momentum 159
3. Moments of Force and Couples 161
4. Kinematics of a Rigid Body 163
5. Three Systems of Coordinates Axes 166
6. The Tensor of Inertia 169
7. Euler's Equations 172
8. The Eulerian Angles 173
9. The Kinetic Energy of a Rotating Body 175
10. Rotation about a Fixed Axis 177
11. Free Rotation of a Rigid Body 181
12. Rotation of a Symmetrical Body about a Fixed Point 184

CHAPTER X. THERMODYNAMICS 185

1. The Problems and Methods of Thermodynamics 186
2. The First Law of Thermodynamics 193
3. The Second Law of Thermodynamics 201
4. Application of the Two Laws 207

CHAPTER XI. STATISTICAL MECHANICS 213

1. The Phase Space 213
2. Distribution in Phase and Liouville's Theorem 215
3. Statistical Equilibrium and the Canonical Distribution 218
4. The Fundamental Assumption of Statistical Mechanics as Applied to Thermodynamics 221
5. Thermodynamic Analogies 224
6. The Phase Integral 227

CHAPTER XII. THE VECTOR FIELD 230

1. The Gradient 230
2. The Divergence 232
3. The Curl 233
4. The Line Integral 234
5. The Surface Integral 235
6. Gauss's Theorem 236
7. Stokes's Theorem 238
8. Tensor Fields 239
9. Orthogonal Curvilinear Coordinates 240
10. Vector Identities 244
11. The Potential 244

CHAPTER XIII. ELECTROSTATICS 246

1. Electrostatic Fields Due to Fixed Charges 246
2. The Effect of an Electrostatic Field on Material Bodies 263

3. The General Electrostatic Distribution Problem 268
4. Energy of an Electrostatic System 278

CHAPTER XIV. MAGNETOSTATICS AND THE INTERACTION OF STEADY CURRENTS 283

1. Ohm's Law and Steady Currents 284
2. Forces between Steady Currents 288
3. Properties of the Vector Potential and Magnetic Induction When Due to Steady Currents Only 294
4. Magnetic Fields Due to Magnetization 296
5. Effect of a Magnetic Field on Material Bodies 302
6. Energy in a Magnetic Field 303

CHAPTER XV. THE ELECTROMAGNETIC FIELD 306

1. The Electrostatic Field 306
2. The Magnetostatic Field 309
3. Electromagnetic Induction 310
4. Fields in Moving Coordinate Systems 313
5. The Energy in a Magnetic Field 318
6. The Displacement Current 323
7. Maxwell's Equations 324
8. Units 325
9. Maxwell's Equations in Moving Systems 326
10. Electromagnetic Field Energy 328
11. Electromagnetic Field Momentum 330
12. General Electromagnetic Potentials 332
13. Electromagnetic Waves in Homogeneous Uncharged Dielectrics 334
14. Lorentz's Form of Maxwell's Equations 335
15. Radiation from an Oscillating Dipole 336

CHAPTER XVI. THE RESTRICTED THEORY OF RELATIVITY . 339

1. Invariance of Newton's Laws under Galilean Transformations 339
2. The Postulates of Relativity 340
3. The Lorentz Transformation 342
4. Transformation of Maxwell's Equations 345
5. Consequences of the Lorentz Transformation for Mechanics 349

INDEX 357

CHAPTER I

ELEMENTARY DIFFERENTIAL EQUATIONS

Most laws of physics are best expressed in the form of differential equations. When Galileo made his studies of falling bodies, he found there was one property common to all the motions. He found that the rate of change of the downward velocity was always the same. The recognition of this common and constant element constitutes the discovery of a law of falling bodies.

To express such a law precisely and simply requires the notation of the differential calculus. Velocity itself is a derivative—the ratio of an infinitesimal distance to the time taken by a body to move through that distance. The rate of change of the velocity is then a second derivative, and the law of falling bodies near the surface of the earth can be written

$$\frac{d^2z}{dt^2} = -g$$

In this equation z is the height of the body, t is the time, and g is a constant known as the acceleration of gravity. This statement of the law is a differential equation. The differential equation is true no matter with what velocity or from what position the body starts to move. As is true with most laws of physics, this law applies only under suitable restrictions. In this case the law is valid when the resistance of the air and the curvature of the earth's surface can be neglected. Under these restrictions the differential equation states a property that is common to the paths of all falling bodies.

Because of the important place that differential equations occupy in mathematical physics, it is necessary to spend a little time in the study of the more elementary methods for finding their solutions. The general study of differential equations is an extensive branch of mathematics and one about which a

physicist cannot know too much. Here, however, only practical questions will be treated, and methods will be given for the solution of a few of the simpler and more common types of equations.

1. Nature of a Differential Equation and Its Solution.—An ordinary differential equation of the first order and the first degree states a functional relationship between a single independent variable x, a dependent variable y, and the derivative dy/dx, in which the derivative appears to the first power only. Such equations can be written in the form

$$\frac{dy}{dx} = f(x,y) \tag{1-1}$$

If y is represented as a function of x by a curve in the x-y plane, equation (1-1) gives the slope of the curve at every point in this plane. The object in solving such a differential equation is to find a relationship between y and x such that equation (1-1) will be satisfied for all values of the independent variable. Geometrically, finding a solution is finding a curve whose slope and coordinates at each point of the curve satisfy equation (1-1).

As an illustration, consider the equation

$$\frac{dy}{dx} = xy \tag{1-2}$$

This says that at every point in the x-y plane the slope of the solution is equal to xy. The significance of this statement is illustrated in Fig. 1-1, where the slope is indicated at a number of points in the plane. In particular, the slope is zero at all points on the x and y axes. At any point on the hyperbola $xy = 1$, the slope is 1; at any point on $xy = -1$, it is -1. Any curves representing solutions that pass through these points must pass through them with the indicated slopes.

It soon becomes apparent from a study of Fig. 1-1 that there is no one single-valued function of x which can be regarded as *the* solution of the differential equation. There is an infinity of such functions. For example, the x axis, $y = 0$, is a solution, for it has everywhere the slope zero. Another

solution is $y = e^{x^2/2}$. This crosses the y axis at $y = 1$ and at this point has a zero slope. In fact, its slope at any point along it is $dy/dx = xe^{x^2/2} = xy$, so that the differential equation is satisfied.

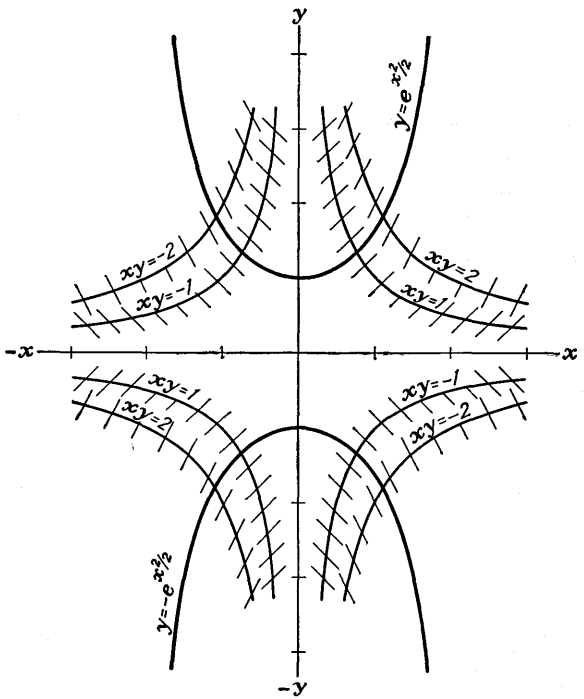


FIG. 1-1.—Some solutions and some lines of constant slope defined by the differential equation $dy/dx = xy$.

It is possible to express this infinity of solutions by the use of an arbitrary constant. Thus it may be said that

$$y = Ae^{x^2/2} \tag{1-2a}$$

is the general solution of equation (1-2), where A is an arbitrary constant. This is meant to imply that equation (1-2a) has the property expressed by equation (1-2) no matter what value is given to A. When $A = 0$, the solution $y = 0$ results. When $A = +1$ or -1, the two solutions illustrated in Fig. 1-1 are obtained.

It is shown in the general theory of differential equations that the general solution of a first-order differential equation will

always contain one arbitrary constant and that a solution containing an arbitrary constant is the general solution.

It is important to recall the significance of the function notation, $f(x,y)$. This means that, if x and y are given specific values, it is possible to find the corresponding value for the function $f(x,y)$. It has nothing to do with the possibility of writing the function in a simple form. This possibility is largely a matter of notation. A function may well be represented by a table of values in which there corresponds to each set of values of the independent variables a definite value of the function. It is always necessary to keep clearly in mind the difference between the existence of a function and the possibility of writing an expression for it.

It is a fundamental problem in the study of a differential equation to find out whether or not a solution really exists. The differential equations of physics represent, in most cases, the results of abstraction from experimental data and hence really the results of differentiating the solutions. For this reason, one is justified, in elementary work, in assuming that a solution exists and in merely undertaking to find it. Furthermore, one may say quite generally that, if in an equation of the type of (1-1) the function $f(x,y)$ is single-valued and continuous, and if it has an absolute value less than a certain upper bound at every point of the x-y plane, a solution does exist.

If a differential equation contains derivatives higher than the first, the equation is said to have the order of the highest derivative. If the highest-order derivative appears to a power higher than the first, the equation is said to be of the degree of this power. The simplest equations, those of the first order and the first degree, will be treated in this chapter.

If the dependent variable is a function of two or more independent variables, a differential equation will contain the partial derivatives with respect to the independent variables. Such an equation is called a *partial differential equation.* Partial differential equations are of great importance in physics, but their solution requires more or less special methods for each type of equation, and therefore the methods of solution will be discussed in connection with the equations of physics themselves.

2. Solution by the Separation of Variables.—If an equation can be written in the form

$$f(x)dx = g(y)dy \tag{1-3}$$

the solution is immediately evident in the form

$$\int f(x)dx = \int g(y)dy + C \tag{1-3a}$$

where C is the arbitrary constant. This method is called *solution by the separation of variables* and is one of the most commonly used.

As an example consider again equation (1-2). If both sides are multiplied by dx and divided by y, there results

$$\frac{dy}{y} = x\,dx \tag{1-3b}$$

This is of the form of equation (1-3), since the left-hand side is a function of y only and the right-hand side is a function of x only. If now both sides are integrated, the result is

$$\log y = \frac{x^2}{2} + C \tag{1-3c}$$

This is the general solution of the equation, and it can be put in the form of equation (1-2*a*) by transforming to the exponential form and defining A as e^C.

Problem 1. For the following differential equations sketch roughly the slope at a number of points in the x-y plane, solve the equations by separation of the variables, and sketch roughly some of the solutions:

a. $\dfrac{dy}{dx} = 2xy$

b. $y\,dx - x\,dy = 0$

c. $\dfrac{dy}{dx} = \dfrac{1}{x \tan y}$

d. $\dfrac{dy}{dx} = -\left(\dfrac{2}{a}\right) xe^{ay}$

3. Formulation of the Differential Equation.—Probably the most difficult part of a problem in mathematical physics is the formulation of the differential equation whose solution gives the solution of the problem. It is necessary to translate the physical statement of the situation into a mathematical state-

ment. The mathematical statement is usually a differential equation.

Skill in the mathematical formulation of physical problems comes largely from practice and experience. It requires a thorough appreciation of the significance of mathematical notation and an intimate and precise understanding of the physical laws involved. It requires that the essential features be recognized and the nonessential features neglected.

As an illustration of the process of writing a differential equation, consider the geometrical problem of finding the curve whose slope at every point is equal to a constant multiple of the abscissa of the point. The first step is to establish a notation. Let $y(x)$ be the function whose curve in the x-y plane has the desired properties, and let k be the constant by which the abscissa is to be multiplied to get the slope. It is then necessary to know, and to remember, that the slope of a curve at any point is the value of the derivative, dy/dx, at that point. The description of the curve then states the equality of this derivative to k times the x coordinate of the point. In equation form,

$$\frac{dy}{dx} = kx \tag{1-4}$$

The solution of this equation is $y = (kx^2/2) + C$. This form of y as a function of x satisfies the differential equation, and the statement of the problem, for any value of C. There is, not one curve, but a whole family of curves that satisfy the condition imposed. It is therefore possible to apply some other condition in addition to that represented by the differential equation. It might be desired to have the curve pass through the point $(x = 0, y = 0)$. This can be accomplished by selecting that member of the family for which $C = 0$.

Problem 2. Find the curve for which the projection of the ordinate on the normal is constant.

Problem 3. Find the curve $y(x)$ whose slope at every point is equal to the negative of the cotangent of the angle between the radius vector and the x axis.

In the case of physical problems, a knowledge of the essential

physical laws and principles must be used to get the proper equation. Consider the problem of finding the atmospheric pressure as a function of the height above the surface of the earth. To establish the notation, consider a vertical column of air of unit cross section, and let the height of a point in this column be z. The problem is then to find the pressure p as a function of z.

Consider a flat element of volume, of thickness dz and unit cross section. The pressure on the top of this element will be less than the pressure on the bottom by the weight of the air in the element. This weight is the mass density ρ multiplied by the volume dz and the acceleration of gravity g. Then, if dp is the amount by which the pressure on the top exceeds the pressure on the bottom,

$$dp = -\rho g\, dz \qquad \text{or} \qquad \frac{dp}{dz} = -\rho g \tag{1-5}$$

This is a differential equation that describes the physical situation, but it is unsatisfactory because it contains too many different quantities whose relationships to each other are not precisely specified. If ρ and g are known as functions of z, these functions can be inserted and the integration carried out. Otherwise, recourse must be had to other information.

To simplify the problem, consider the fact that the value of g does not change very much in heights to which the atmosphere extends. For this reason a very close approximation can be obtained by treating g as a constant of the equation. From the laws of a perfect gas it is known that a close approximation is obtained by setting the density proportional to the pressure, if the temperature is constant. The temperature is really not constant; but if this approximation is made,

$$\frac{dp}{dz} = -Kgp \tag{1-6}$$

where K is the assumed constant of proportionality. The solution of this equation is

$$p = Ae^{-Kgz} \tag{1-7}$$

and A can be adjusted to give the correct value of the pressure at $z = 0$ or some other point.

If it is not desired to make the approximation of constant temperature, other information must be used, such as the value of the temperature as a function of z. In any case, it is necessary to get the equation into such a form that it contains p and z and the derivative only.

Problem 4. A rod of circular cross section is to support a tension T when in a vertical position. If the density of the rod is ρ, find the shape it must have in order that the tension per unit cross section shall everywhere be τ.

4. Linear Differential Equations.—An equation of the form

$$\frac{dy}{dx} + Py = Q \tag{1-8}$$

is said to be a linear differential equation if P and Q are not functions of y. They may be constants or any functions of x. Linear equations, especially those of the second order, are of particular importance in physics, and considerable attention will be given to them later. For the first-order equation a general solution can be written in the form

$$y = e^{-\int P\,dx}\left(\int Qe^{\int dx}\,dx + C\right) \tag{1-9}$$

To show that this is a solution, it is necessary only to differentiate y with respect to x and show that the derivative satisfies equation (1-8). The presence of the arbitrary constant C makes the solution general.

As an example of a linear equation take

$$\frac{dy}{dx} + \frac{y}{x} = x^2 \tag{1-10}$$

Comparison with equation (1-8) shows that $P = 1/x$ and $Q = x^2$. Substitution in (1-9) then gives

$$\begin{aligned} y &= e^{-\int dx/x}\left(\int x^2 e^{\int dx/x}\,dx + C\right) \\ &= \frac{1}{x}\left(\int x^3\,dx + C\right) = \frac{x^3}{4} + \frac{C}{x} \end{aligned} \tag{1-11}$$

Problem 5. Find the solution of the linear equation

$$\frac{dy}{dx} + 2xy = 2e^{-x^2}$$

that passes through the point $x = 0, y = 1$.

Problem 6. Find the solution of

$$\frac{dy}{dx} + y \sin x = A \sin x$$

that passes through $x = 0, y = 1$.

Problem 7. Find the solution of

$$\frac{1}{x^2}\frac{dy}{dx} + y = A$$

that passes through $x = 1, y = 1$.

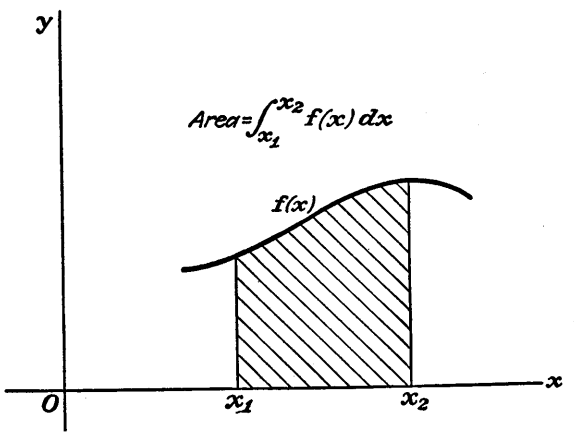


FIG. 1-2.—An illustration of the geometrical representation of a definite integral by means of an area.

5. Definite and Indefinite Integrals.—Most methods for finding the solution of a differential equation give the result in the form of an integral. The evaluation of such an integral in terms of simple functions may be difficult or impossible, but this is largely a matter of notation. The significance of the integral is quite clear in any case, since it can always be evaluated to any desired degree of accuracy by graphical or numerical methods.

A useful definition of the integral of a function $f(x)$ between two limits x_1 and x_2 is that it is the area bounded by the curve representing $f(x)$, the x axis, and the straight lines perpendicular to the x axis at x_1 and x_2. This is shown in Fig. 1-2. It can be

evaluated, if necessary, by plotting the curve and measuring the proper area. Such a definite integral is a number; it is not a function of x.

The integrals that appear in the solutions of differential equations are not, however, usually definite integrals. They are indefinite integrals. An indefinite integral can be defined as a function whose derivative is the integrand in question. It is often convenient, however, to define it in the same way as a definite integral by means of an area. If the definite integral is evaluated for different values of its upper limit, it can be regarded as a function of the upper limit. Thus

$$\int f(x)dx + C = \int_{u_0}^{x} f(u)du + C' \tag{1-12}$$

The difference between the constants C and C' is connected with the lower limit u_0 from which the integration is carried out, and the variable u is used under the integral sign on the right-hand side to distinguish clearly between the variable of integration and the limit. On this basis an indefinite integral can be evaluated graphically or numerically by evaluating a definite integral for different values of the upper limit and then plotting or tabulating these values as a function of this upper limit.

Problem 8. Let $f(x)$ be such that it is equal to zero for $x < 0$, is equal to x for $0 < x < a$, and is equal to a for $x > a$. Evaluate the indefinite integral

$$I = \int f(x)dx$$

Problem 9. Integrate $f(x)$ when $f(x) = \sin x$ for $0 < x < \pi$ and is zero elsewhere.

6. Simpson's Rule and Numerical Integration.—It is usually possible to evaluate integrals numerically with greater accuracy and less labor than is involved in graphical methods. One numerical method, called *Simpson's rule,* is based on the idea of approximating small portions of the integrand by a quadratic curve and then integrating the quadratic function analytic-

ally. Consider the situation indicated in Fig. 1-3 when it is desired to integrate the function $f(x)$ between the limits x_0 and x_2. The numerical procedure is to replace $f(x)$ by the polynomial $a_0 + a_1x + a_2x^2$ that equals the given function $f(x)$ for x_0, x_1, and x_2. x_1 is selected so that $x_1 - x_0 = x_2 - x_1$. The approximation involved is in the replacement of $f(x)$ by this polynomial. The polynomial can be directly integrated.

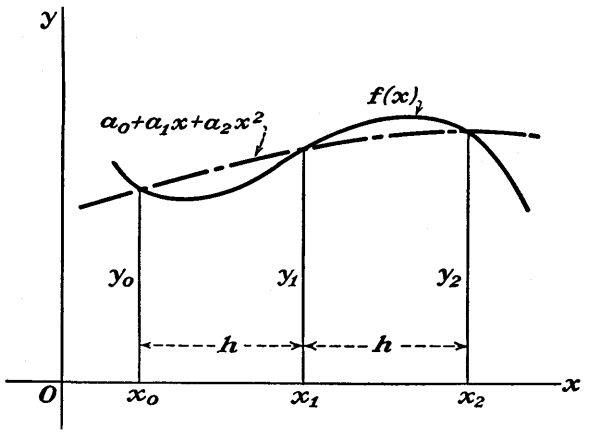


FIG. 1-3.—Approximation to an arbitrary curve by means of a quadratic curve, as used in Simpson's rule for numerical integration.

To determine the coefficients in the polynomial, it is convenient to move the origin of x to x_0 and then to write the three equations that represent the intersection of the polynomial curve with the curve $f(x)$ at the three points (x_0,y_0), (x_1,y_1), (x_2,y_2). These are

$$\left.\begin{aligned} a_0 &= y_0 \\ a_0 + a_1h + a_2h^2 &= y_1 \\ a_0 + a_12h + a_24h^2 &= y_2 \end{aligned}\right\} \tag{1-13}$$

From these it follows that

$$\left.\begin{aligned} a_0 &= y_0 \\ a_1 &= -\frac{1}{h}\left(\frac{3}{2}y_0 - 2y_1 + \frac{1}{2}y_2\right) \\ a_2 &= \frac{1}{2h^2}(y_2 - 2y_1 + y_0) \end{aligned}\right\} \tag{1-13a}$$

The integral of the polynomial carried out between the limits 0 and $2h$ is

$$I = 2ha_0 + 2h^2a_1 + \frac{8h^3}{3}a_2$$
$$= \frac{h}{3}(y_0 + 4y_1 + y_2) \qquad (1\text{-}13b)$$

This expression is entirely independent of the position of the origin. It depends only on the three values of the function y_0, y_1, and y_2 and the distance h between the corresponding values of x.

As indicated above, the only approximation involved in this method of integration is the replacement of $f(x)$ by the quadratic expression. This will be more exact the shorter the range of integration, or the smaller the value of h. Hence, to get a good approximation to the desired integral, it is possible to divide the region of integration into an even number of regions of length h and to integrate each pair of them by the above process. The result of adding these all together is then

$$I = \frac{h}{3}(y_0 + 4y_1 + 2y_2 + 4y_3 + 2y_4 + 4y_5 + \cdots + y_{2n}) \qquad (1\text{-}14)$$

This is the usual form of Simpson's rule.

Problem 10. Use Simpson's rule to evaluate the integral

$$\int_{0.2a}^{1.8a} (2ax - x^2)^{-\frac{1}{2}}\, dx$$

Divide the range of integration into four equal parts, and compare the result with the result of analytic integration.

Problem 11. Integrate the $\sin x$ from $x = 0$ to $x = \pi$ by Simpson's rule. First divide the range of integration into two intervals and then into four, and compare the results with the exact value of the integral.

Problem 12. Work out a formula analogous to equation (1-13b) by dividing the range of integration into three equal parts and approximating the function by a cubic expression.

Problem 13. Evaluate the integral

$$\int_0^2 \frac{\sin x^2}{x}\, dx$$

numerically.

7. Differentiation of Integrals.—A definite integral is a function of the limits of integration and of any parameters that may be in the integrand; therefore it can be differentiated with respect to any of these quantities. It is not, however, a function of the variable with respect to which the integration has been carried out. Since it is occasionally necessary to differentiate integrals, the rules will be written down here without any further discussion. They are almost obvious from the definition of an integral.

Differentiation with respect to the upper limit:

$$\frac{\partial}{\partial b}\int_a^b f(x,t)dx = f(b,t) \qquad (1\text{-}15a)$$

Differentiation with respect to the lower limit:

$$\frac{\partial}{\partial a}\int_a^b f(x,t)dx = -f(a,t) \qquad (1\text{-}15b)$$

Differentiation with respect to a parameter:

$$\frac{\partial}{\partial t}\int_a^b f(x,t)dx = \int_a^b \frac{\partial f(x,t)}{\partial t}\, dx \qquad (1\text{-}15c)$$

Problem 14. Illustrate the use of equation (1-15*c*) by integrating $\sin tx$ between $x = a$ and $x = b$ and then differentiating the integral. Show that the result is the same as that obtained by first differentiating with respect to t and then integrating with respect to x.

Problem 15. The density of a thin straight rod varies as $e^{\alpha x}/x$, where x is measured along the rod. One end of the rod is at $x = a$ and the other end at $x = b$. Both a and b are positive. If the cross section of the rod is constant, find the rate of change of the total mass of the rod with a change in α.

8. Second-order Equations That Are Reducible to the First Order.—Some second-order equations can be reduced to the

first order by considering dy/dx as a new variable p. Among these are two important classes.

a. Equations That Do Not Contain y.—These are of the form

$$f\left(\frac{d^2y}{dx^2}, \frac{dy}{dx}, x\right) = 0 \tag{1-16}$$

Upon substitution of p for dy/dx this immediately becomes a first-order equation,

$$f\left(\frac{dp}{dx}, p, x\right) = 0 \tag{1-16a}$$

If this can be solved to give p as a function of x, the result is a first-order differential equation whose solution gives y as a function of x.

As an example, consider the equation

$$\frac{d^2y}{dx^2} - x \sin x = 0 \tag{1-17}$$

Substitution of p for dy/dx gives immediately the first-order equation

$$\frac{dp}{dx} = x \sin x \tag{1-17a}$$

This equation is separable, and its general solution is

$$p = A - x \cos x + \sin x = \frac{dy}{dx} \tag{1-17b}$$

where A is the arbitrary constant. This again is a first-order equation that is separable; therefore its general solution can be written down.

$$y = B + Ax - x \sin x - 2 \cos x \tag{1-17c}$$

This solution has two arbitrary constants, A and B, and both of them are necessary to give a general solution. In the same way in which a first-order equation will have one arbitrary constant in its general solution, a second-order equation will have two such constants. To select any one function as a solution of the equation, it is necessary to have

two additional specifications. These may be that the curve passes through two specified points, that it passes through some one specified point with a specified slope, or some other pair of similar conditions.

Problem 16. Find the solution of equation (1-17) that passes through the points $(x = 0, y = 0)$, and $(x = \pi, y = 0)$.

Problem 17. Find the solution of equation (1-17) that passes through the origin with a slope of -1.

Problem 18. Find the solution of

$$\frac{d^2y}{dx^2} = x^m (m > 0) \tag{1-18}$$

that passes through $(x = 1, y = 1)$ with zero slope.

Problem 19. Find the general solution of

$$\frac{d^2y}{dx^2} - \left[a^2 + \left(\frac{dy}{dx}\right)^2\right]^{\frac{1}{2}} = 0 \tag{1-19}$$

b. Equations That Do Not Contain x.—These are of the form

$$f\left(\frac{d^2y}{dx^2}, \frac{dy}{dx}, y\right) = 0 \tag{1-20}$$

A simple substitution of p for dy/dx does not suffice in this case; but if it is accompanied by a transformation that makes y the independent variable, a solution can be sought that gives p as a function of y. This can be done by making use of the fact that

$$\frac{d^2y}{dx^2} = \frac{dp}{dx} = \frac{dp}{dy}\frac{dy}{dx} = p\frac{dp}{dy}$$

Then equation (1-20) becomes

$$f\left(p\frac{dp}{dy}, p, y\right) = 0 \tag{1-20a}$$

which is a first-order differential equation with p as the dependent variable and y as the independent variable. The general solution of this equation will give p as a function of y and an arbitrary constant. This again is a first-order differential

equation, whose solution gives y as a function of x and a second arbitrary constant. As an example, consider the equation

$$\frac{d^2y}{dx^2} + \frac{1}{y^3} = 0 \tag{1-21}$$

The substitution of p for dy/dx leads to

$$p\frac{dp}{dy} + \frac{1}{y^3} = 0 \tag{1-21a}$$

This is now a first-order equation that is separable, and the general solution is

$$p^2 - \frac{1}{y^2} = A$$

This result can be transformed to

$$\frac{dy}{dx} = \left(A + \frac{1}{y^2}\right)^{\frac{1}{2}} \tag{1-21b}$$

which is again separable and has as its general solution

$$y = \left[A(x + B)^2 - \frac{1}{A}\right]^{\frac{1}{2}} \tag{1-21c}$$

The member of this family of curves that passes through the point $(x = 0, y = 1)$ with zero slope has $B = 0, A = -1$ so that $y = (1 - x^2)^{\frac{1}{2}}$.

Problem 20. Find the solution of

$$y\frac{d^2y}{dx^2} + \left(\frac{dy}{dx}\right)^2 + 1 = 0$$

that passes through the origin with zero slope.

Problem 21. Find a curve such that its tangent at the point (x,y) crosses the y axis at a distance x above the origin.

Problem 22. Find the general solution of

$$\frac{d^2y}{dx^2} - \alpha y = 0$$

when α is positive.

Problem 23. Show that, if y satisfies the equation

$$\frac{d^2y}{dx^2} + Iy = 0$$

where I is a function of x only, the quantity $u = y^2$ satisfies the equation

$$\frac{d^3u}{dx^3} + 4I\frac{du}{dx} + 2\frac{dI}{dx}u = 0$$

Problem 24. Find the general solution of

$$x\frac{d^2y}{dx^2} - \frac{dy}{dx} = 0$$

and interpret the arbitrary constants in terms of the curves representing y as a function of x.

Problem 25. Find a curve passing through the origin such that the projection of the normal on the x axis is equal to $\sin x$.

CHAPTER II

THE MECHANICS OF PARTICLES

In the study of the mechanics of particles, we shall be concerned almost exclusively with the derivation of consequences of the fundamental laws. In the early stages of the development of a science, interest centers principally in the study of experimental results for the purpose of discovering in them the laws that can be formulated. The science of ordinary mechanics passed this stage more than two centuries ago. There is now no question but that the behavior of ordinary material particles and rigid bodies, moving with velocities small compared with the velocity of light, is correctly described by Newton's laws of motion. The interest now lies in the purely mathematical problem of finding the consequences of these laws in a variety of cases and in finding the forces that act under different circumstances.

It is possible, of course, to state the fundamental laws of mechanics in other forms than that selected by Newton. Hamilton's principle, which is one of these other forms, will be treated in a later chapter. The original form, however, appeals so directly to one's sense of physical reality that there seems to be good reason for using it as the basis from which the others can be derived.

1. Newton's Equations of Motion.—Newton's first and second laws of motion for a system of n particles can be expressed by means of $3n$ differential equations of the second order. Let the mass and the rectangular coordinates of the ith particle be m_i and x_i, y_i, z_i, and let the three corresponding components of the total force on the ith particle be X_i, Y_i, Z_i.

In this book the coordinates x, y, z will always refer to a right-hand system of Cartesian coordinates such as is shown in Fig. 2-1. This specification is of some importance in the formula for vector multiplication and in the electromagnetic

equations, where a change from a right-hand to a left-hand system involves a change in sign.

Newton's equations of motion for the ith particle are then

$$\frac{d}{dt}\left(m_i \frac{dx_i}{dt}\right) = X_i \qquad \frac{d}{dt}\left(m_i \frac{dy_i}{dt}\right) = Y_i \qquad \frac{d}{dt}\left(m_i \frac{dz_i}{dt}\right) = Z_i \quad (2\text{-}1)$$

For the cases to which these equations will be applied, the mass is a constant, and the equations can be written

$$m_i \frac{d^2x_i}{dt^2} = X_i \qquad m_i \frac{d^2y_i}{dt^2} = Y_i \qquad m_i \frac{d^2z_i}{dt^2} = Z_i \quad (2\text{-}2)$$

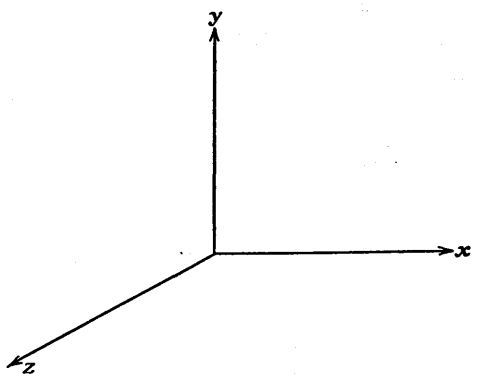


FIG. 2-1.—A right-hand system of rectangular coordinate axes.

These are equations of the second order and must be treated by methods suitable for such equations. In some cases they can be reduced to equations of the first order and solved by the methods of the previous chapter. The solutions are expressions that give the positions of the particles as functions of the time and the $6n$ arbitrary constants of integration. These arbitrary constants can then be determined from the known values of the positions and velocities at any given time. Equations (2-1) and (2-2) are generally known as Newton's equations, although much of their content was developed by Galileo.

The confidence in the validity of these equations is based upon the correctness of their solutions as descriptions of observed motions. It is important to remember, however, that their use involves the ability to define the quantities contained in

them. The idea of accleration can be made precise in terms of ideas of position and time. These must be assumed. The ideas of mass and force are less common, and Newton's attempts at definitions of them were not very satisfactory. It is not difficult to get an intuitive idea of force in terms of muscular sensations, although it is not so easy to get the precise significance of the term as used in mechanics. The variety of the possible ideas is well illustrated by the wide range of concepts covered by the term *force* in popular language. It is also possible to get an intuitive idea of mass, although again it is not so easy to distinguish it from bulk or weight. These definitions will not be discussed here, since it is assumed that the reader has an adequate understanding of them. For further analysis, reference may be made to the work of Mach.*

2. The Energy Integral.—For a complete solution of a mechanical problem, it is necessary to know the forces on each particle, X_i, Y_i, Z_i, as functions of the positions of all the particles and of the time. It is the knowledge of these forces that constitutes the knowledge of the physical laws governing the situation. Because of Newton's equations, a statement of the forces is equivalent to a statement of the accelerations; and the accelerations at all times together with the initial positions and velocities determine the positions at any later time. It is possible, however, without a complete knowledge of the forces, to obtain some integrals of the equations that depend on very general properties of the forces. Such integrals are often of considerable assistance in the solution of a problem.

If in equations (2-2) one writes $\dot{x}_i$ for dx_i/dt and multiplies both sides of the equation by $\dot{x}_i$, the result is

$$m_i \frac{d\dot{x}_i}{dt} \dot{x}_i = X_i \dot{x}_i = X_i \frac{dx_i}{dt} \tag{2-3}$$

Multiplication by dt and integration of both sides gives

$$\tfrac{1}{2} m_i \dot{x}_i^2 = \int X_i \, dx_i + C_i \tag{2-4}$$

* MACH, E., "Science of Mechanics," p. 238, The Open Court Publishing Company, LaSalle, Ill.

The quantity $X_i\,dx_i$ is, by definition, the work done by the component of force X_i when the particle is displaced by the amount dx_i. Correspondingly, the integral is the amount of work done by X_i when the particle moves between the limits of integration. This, however, is not a very useful quantity, for only in especially simple cases is the value of the integral independent of the particular path over which the particle is moved.

In case the x component of the force, X_i, is a function of the coordinate x_i only, the integral $\int X_i\,dx_i$ can be evaluated and expressed in terms of the values of x_i at the limits only. Such cases occur in some simple problems, such as that of a projectile subject to a constant downward force. If the x axis is taken as pointing vertically upward, $\int X\,dx = -mg(x - x_0)$ where m is the mass of the particle, g the acceleration of gravity, and x_0 is an arbitrary constant or lower limit of integration. The negative of this quantity, $mg(x - x_0)$, may be called the potential energy in the x direction, when the particle has the coordinate x. If it is designated by V_x, equation (2-4) leads to

$$T_x + V_x = C \tag{2-5}$$

where $T_x = (m/2)\dot{x}^2$ is the kinetic energy in the x direction.

It is only in special cases that such a separation of the potential energy into parts along the various axes can be carried out. More frequently, each component of the force depends on all the coordinates of the system, and the work done by any one component of the force depends upon the path along which the particle is moved. For example, consider the case of a single particle attracted toward the origin of coordinates with a force that is inversely proportional to the square of the distance. The force on the particle when it is at the point (x,y,z) is $K/(x^2 + y^2 + z^2)$, directed toward the origin. The x component of the force is then

$$X = \frac{-Kx}{(x^2 + y^2 + z^2)^{\frac{3}{2}}} \tag{2-6}$$

Now let the integral be evaluated when the particle is moved from $(x_0,0,0)$ to $(x_1,0,0)$ along the x axis. The integral can be easily evaluated since y and z are always zero.

$$X = -\frac{K}{x^2} \quad \text{and} \quad -\int_{(x_0,0,0)}^{(x_1,0,0)} \frac{K}{x^2}\,dx = \frac{K}{x_1} - \frac{K}{x_0}$$

However, suppose the particle is moved between these two points along a different path. Suppose it is moved along the

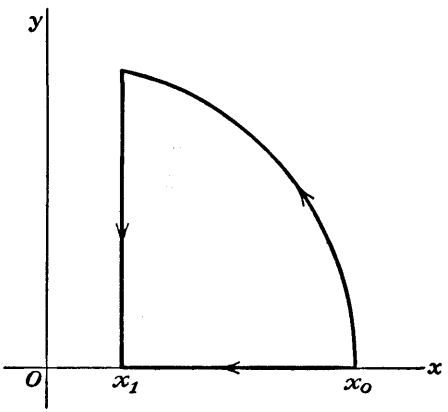


FIG. 2-2.—Two possible paths of integration between $(x_0, 0, 0)$ and $(x_1, 0, 0)$.

path indicated in Fig. 2-2. Along this path the denominator in equation (2-6) is constant from $(x_0,0,0)$ to $(x_1,\sqrt{x_0^2 - x_1^2},0)$, and therefore $X = -Kx/x_0^3$. The integral is then

$$-\frac{1}{x_0^3}\int_{x_0}^{x_1} Kx\,dx = K\left(\frac{x_0^2}{2x_0^3} - \frac{x_1^2}{2x_0^3}\right)$$

over the first part of the path. Along the second part of the path the integral vanishes, since there is no motion in the x direction. It can be seen clearly from this example that the value of such an integral may depend very strongly upon the path over which the particle is moved, so that the integral itself has a very restricted usefulness.

In spite of the fact just illustrated, there are a great many

forces of common occurrence, with such a form that the sum of the set of integrals

$$-\sum_{i=1}^{n}\int(X_i\,dx_i + Y_i\,dy_i + Z_i\,dz_i) = V(x_1,y_1,z_1, \ldots, z_n) \quad (2\text{-}7)$$

does depend on the limits of integration only. This integral can then be given the name, *the potential energy of the system.*

The sum of the terms containing the squares of the velocity components is called *the kinetic energy of the system* T, and the sum of equations (2-4) gives

$$\sum_{i=1}^{n}\left[\frac{m_i}{2}(\dot{x}_i^2 + \dot{y}_i^2 + \dot{z}_i^2) - \int (X_i\,dx_i + Y_i\,dy_i + Z_i\,dz_i)\right] = T + V = C$$

This is the theorem of the conservation of energy. It states that the sum of the quantity defined as T and the quantity defined as V does not change with the time but is a constant throughout the motion.

There are several points in connection with the law of conservation of energy that are worthy of mention and emphasis.

1. The law is in one sense a description of a certain class of forces, the conservative forces. Only when the forces involved are of this type does the law apply. When this is recognized, it can also be recognized that the quantities defined as energies have been so defined because their sum is a constant. When the kinetic and potential energies are properly defined, the conservation of mechanical energy is a direct consequence of the laws of motion. It is not an additional law, but it is often of great assistance in handling mechanical problems.

2. The potential energy can be defined in two ways. It can be defined, as above, as the negative of the work done by the forces of the system when the particles move from one configuration to another. The final configuration is the one to which the potential energy is assigned. The initial configuration is arbitrarily selected as that for which the potential energy is defined to be zero.

On the other hand, the potential energy can be defined as

that function of the coordinates $V(x_i,y_i,z_i)$ whose negative partial derivatives are the components of the forces.

$$-\frac{\partial V}{\partial x_i} = X_i \tag{2-8}$$

These two definitions are obviously equivalent, for if these values of the forces are inserted in equation (2-7) it becomes an identity.

3. The actual value of the potential energy is entirely arbitrary. Equation (2-8) is unaffected by the addition of any constant to V, and in equation (2-7) the position of the initial configuration can be arbitrarily chosen.

4. The potential energy as defined in equation (2-7) is the negative of the work done by the forces of the system. Another point of view is to regard it as the work done by some outside agent in moving the system, against the forces of the system, from the initial to the final configuration. Since the force exerted by the outside agent must balance and overcome the forces of the system, the change in sign is accounted for.

In the case of the particle subject to the uniform force of gravity, the outside agent must exert an *upward* force of magnitude mg, and the work done by this agent is $mg(x - x_0)$.

Problem 1. Use the energy integral to determine the height to which a projectile will rise when it is thrown upward.

Problem 2. Show by integration along the two paths in Fig. 2-2 that the sum of the work done by all components of the force is a function of the end points of the path only.

3. Newton's Third Law of Motion and the Momentum Integral.—In addition to his first two laws of motion which can be stated in the equations of motion, Newton stated a third law of motion according to which two interacting bodies suffer equal and opposite changes of momentum in each interval of time. This implies that the force on one particle due to a second is equal and opposite to the force on the second due to the first, and that this equality is independent of the velocities of the two particles.

Let X_{ij} be the force on particle i due to particle j. Then

$$X_i = \sum_j X_{ij} + X_i^0 \tag{2-9}$$

The sum in this equation is over all values of j from 1 to n. Since this includes the case of $j = i$, the quantity X_{ii} is defined to be identically zero, since it has no real significance. The quantity X_i^0 is the x component of the force on the ith particle that is not associated with any other particle. It can be regarded as an "outside" force. Newton's third law states that

$$X_{ij} = -X_{ji} \tag{2-10}$$

If the outside forces are zero and the only forces are those between particles, equation (2-1) for the ith particle becomes

$$\frac{d}{dt}\left(m_i \frac{dx_i}{dt}\right) = X_i = \sum_j X_{ij} \tag{2-11}$$

If a similar equation is written for each particle and they are all added together, the result is

$$\frac{d}{dt}\sum_i m_i \frac{dx_i}{dt} = \sum_{i,j} X_{ij} = 0 \tag{2-12}$$

The double sum is over all values of i and j from 1 to n. It is zero because to each X_{ij} must be added X_{ji}, and the sum of this pair of quantities is zero by equation (2-10).

The integral of equation (2-12) can be immediately written down. It is

$$\sum_i m_i \frac{dx_i}{dt} = M_x \tag{2-13}$$

The constant M_x is called the x component of the total momentum of the system. The same argument can be applied to the other two directions, and the result is that all three components of the momentum are constants under the conditions specified, *viz.*, that there are no external forces acting. This is called the *theorem of the conservation of momentum.* It is a direct

consequence of the laws of motion, including the third. One might even say that Newton's third law of motion is a statement of the conservation of momentum.

Problem 3. Write out equation (2-12) for the case of three interacting particles, and show that the sum of the forces is zero.

Problem 4. Show that the rate of change of the total momentum is equal to the total external force acting.

Problem 5. The coordinates of the center of mass of a system of particles are, by definition, given by

$$\xi = \frac{\Sigma m_i x_i}{\Sigma m_i}, \qquad \eta = \frac{\Sigma m_i y_i}{\Sigma m_i}, \qquad \zeta = \frac{\Sigma m_i z_i}{\Sigma m_i} \tag{2-14}$$

Show that the center of mass will move in the same way as a particle whose mass is equal to the total mass of the system and which is acted on by the sum of all the forces acting on the particles of the system.

Show that if the only forces are mutual forces between the particles, the center of mass of the system will move as a free particle moves.

4. The Conservation of Angular Momentum.—If the first of equations (2-2) is multiplied by y_i and the second by x_i, the difference of the two equations is

$$m_i\left(x_i \frac{d^2y_i}{dt^2} - y_i \frac{d^2x_i}{dt^2}\right) = x_iY_i - y_iX_i \tag{2-15}$$

The left side of this can be transformed to give

$$m_i \frac{d}{dt}(x_i\dot{y}_i - y_i\dot{x}_i) = x_iY_i - y_iX_i \tag{2-16}$$

The quantity in the parentheses is equal to twice the rate at which a perpendicular from the particle to the z axis sweeps out area over the x-y plane. The right-hand side of (2-16) is called the *moment of force* about the z axis, acting on the ith particle. The quantity $m_i(x_i\dot{y}_i - y_i\dot{x}_i)$ is called the *angular momentum* or, sometimes, the *moment of momentum* of the ith particle about the z axis. Equation (2-16) states the theorem that the rate of change of the angular momentum about an axis is equal to the moment of force about that axis.

If both sides of equation (2-16) are summed for all the

particles, the sum of the left sides gives the rate of change of the total angular momentum of the system and the sum of the right sides is the total moment of force.

Problem 6. Show by a diagram that $x_i\dot{y}_i - y_i\dot{x}_i$ is twice the areal velocity of the ith particle about the z axis.

Problem 7. Show that the moment of force acting on a single particle is equal to the component of force perpendicular to the radius vector to the particle from the origin, multiplied by the length of the radius.

Problem 8. Show that the rate of change of the angular momentum about the x axis is equal to the moment of force about that axis.

Problem 9. Show that, if a single particle is subject to a force that is directed toward or away from the origin of coordinates, the three components of angular momentum will be constants, and the particle will move in a plane.

Problem 10. Show that for a system of three particles subject to no external forces and in which the mutual forces between pairs of particles are directed along the lines connecting them the total angular momentum about any axis is a constant.

Problem 11. Write and integrate the equations of motion for a particle subject to no forces. Evaluate the arbitrary constants in terms of the initial position and velocity, and evaluate the components of angular momentum.

5. The Conservation Laws of Mechanics.—The three conservation laws just discussed are the most widely useful. In special cases, when special kinds of forces are involved, other quantities also may be conserved. Such quantities are not given names that have come into general use, for they represent situations that are too special to be of wide application.

It is worth while to summarize the conservation laws and the conditions under which they apply.

a. The law of conservation of energy is valid when the forces are conservative, *i.e.*, when they can be derived from a potential energy that is a function of the coordinates only. This statement is really a tautology, for it is only in such cases that the potential energy is defined.

b. The law of conservation of momentum is valid when the forces obey Newton's third law of motion. Again this

statement is fairly trivial, for Newton's third law of motion is equivalent to the law of conservation of momentum. The physical significance of the third law of motion is that it does hold in a very large number of cases and can be made to hold in others by suitable definitions.

c. The law of conservation of angular momentum is valid when Newton's third law of motion holds and when the forces between particles are central forces. This is a more severe restriction than is necessary for the conservation of linear momentum, but it nevertheless is satisfied in a great many cases.

6. The Motion of a Projectile.—Consider a particle under the influence of gravity only. Take a system of axes in which the z axis points vertically upward, so that the x and y axes are horizontal. Only the case in which the particle remains so near the surface of the earth that the acceleration of gravity may be considered constant will be treated here. Under these conditions the equations of motion are

$$\frac{d^2x}{dt^2} = 0 \qquad \frac{d^2y}{dt^2} = 0 \qquad \frac{d^2z}{dt^2} = -g \tag{2-17}$$

Although these are equations of the second order, they can be reduced at once to first-order equations of which the integrals are

$$\frac{dx}{dt} = \dot{x}_0 \qquad \frac{dy}{dt} = \dot{y}_0 \qquad \frac{dz}{dt} = -gt + \dot{z}_0 \tag{2-18}$$

where $\dot{x}_0$, $\dot{y}_0$, and $\dot{z}_0$ are constants of integration whose values must be determined by the conditions of the problem.

Equations (2-18) can be integrated again, to give

$$x = \dot{x}_0 t + x_0 \qquad y = \dot{y}_0 t + y_0 \qquad z = -\tfrac{1}{2}gt^2 + \dot{z}_0 t + z_0 \tag{2-19}$$

These equations give the position of the particle as a function of the time. The equations for x and y show that the projection of the path on the x-y plane is a straight line, so that it is possible to take this line as the x axis and to make y always

equal to zero. The solution of the problem is then given by

$$x = \dot{x}_0 t + x_0 \qquad \text{and} \qquad z = -\tfrac{1}{2}gt^2 + \dot{z}_0 t + z_0 \quad (2\text{-}20)$$

Problem 12. Interpret the constants in equations (2-19) and (2-20) in terms of the initial position and velocity of the particle.

Problem 13. Show from equations (2-19) that the projection of the path on the x-y plane is a straight line.

Problem 14. Find the transformation of coordinates x, y, z, in equations (2-19), to x', y', z' such that y' will be identically zero.

Problem 15. Eliminate t between equations (2-20) to find the trajectory of the particle.

Problem 16. Find the constants of integration in equations (2-20) in terms of the initial position, speed, and inclination of the path to the horizontal.

Problem 17. If a particle starts from the origin of coordinates, with a speed v, along a path that makes the angle α with the horizontal, at what distance from the starting point will it cross a line, through the origin, that makes the angle β with the horizontal?

Problem 18. Show that the law of conservation of energy applies to the motion of a projectile.

Problem 19. Compute the components of momentum and of angular momentum for a projectile, and show that their rates of change are given by equations (2-11) and (2-16).

If it is desired to take into account the resistance of the air through which the particle moves, it is necessary to include in the differential equations of motion some terms to represent the resisting force. The exact law of force is complicated, but for very low velocities the resistance may be taken as proportional to the first power of the velocity. Let the constant of proportionality be R, and let the motion be in the x-z plane. This restriction of the motion to a plane does not affect the generality of the result, since it can be shown that the motion will always be in a vertical plane. The equations of motion are

$$\frac{d^2x}{dt^2} = -\frac{R}{m}\frac{dx}{dt} \qquad \text{and} \qquad \frac{d^2z}{dt^2} = -\frac{R}{m}\frac{dz}{dt} - g \quad (2\text{-}21)$$

These equations can be reduced to those of the first order by

the methods of the previous chapter and can then be solved by the separation of the variables. The solutions are

$$x = -\frac{mA}{R}e^{-Rt/m} + B \qquad \text{and}$$

$$z = -\frac{mC}{R}e^{-Rt/m} - \frac{mgt}{R} + D \tag{2-22}$$

The quantities A, B, C, and D are the constants of integration, which must be determined from the initial conditions.

Problem 20. Find the solutions of equations (2-21) by the methods of Chap. I.

Problem 21. Evaluate the constants in equations (2-22) in terms of the initial position and velocity of the particle.

Problem 22. Show that the velocity of a particle falling under gravity and against a resistance proportional to the velocity tends to approach a constant value as the time increases.

7. Simple-harmonic Motion.—One very common type of motion is that in which a particle is attracted toward a position of equilibrium with a force proportional to the displacement from it. When the motion is confined to a straight line, the equation of motion is

$$m\frac{d^2x}{dt^2} = -ax \tag{2-23}$$

This may be integrated by the methods of Chap. I and gives

$$x = A\sin(\omega t + \delta) \qquad \text{or} \qquad x = A\cos(\omega t + \delta') \tag{2-24}$$

These two results are essentially the same, since one can be transformed into the other by setting $\delta = \delta' + (\pi/2)$. A and δ are the arbitrary constants of integration, while $\omega = \sqrt{a/m}$ and thus is determined by the differential equation itself.

It is often convenient to write equation (2-24) in a slightly different form. If the sine of the sum is expanded, the expression for x becomes

$$x = A\sin\omega t\cos\delta + A\cos\omega t\sin\delta$$

whence

$$x = C\sin\omega t + D\cos\omega t \tag{2-25}$$

Problem 23. Integrate equation (2-23) by reduction to an equation of the first order and the separation of variables.

Problem 24. Evaluate the constants in equation (2-24) in terms of the initial position and velocity of the particle.

Problem 25. Evaluate the constants in equation (2-25) in terms of the initial position and velocity of the particle.

Problem 26. Show from the definition of potential energy that $ax^2/2$ is the potential energy of a particle attracted toward the origin

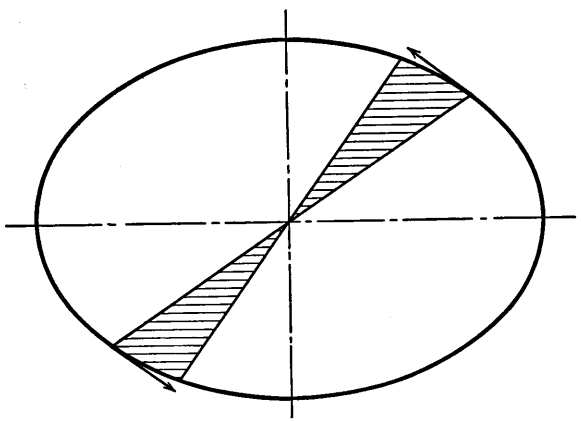


FIG. 2-3.—When a particle is attracted toward a fixed point with a force proportional to the distance from it, the orbit is an ellipse with the attracting point at the center.

of x, with a force equal to $-ax$, when the potential energy at the origin is set equal to zero.

Problem 27. Evaluate the total energy of a particle moving with simple-harmonic motion, in terms of its initial position and velocity.

Problem 28. Evaluate the energy of a particle moving with simple-harmonic motion, in terms of the amplitude of vibration and the frequency.

Problem 29. If a particle is permitted to move in a plane and is attracted toward the origin with a force proportional to the distance from it, the motion is called *plane simple harmonic.* Write and solve the differential equations for this case. Eliminate t between the two solutions to get the path of the particle, and show that this is in general an ellipse (Fig. 2-3). Find the conditions under which it is a circle.

Problem 30. Show that, if in plane simple-harmonic motion the velocity at any time is equal to zero, the motion is in a straight line.

Problem 31. Show from the theorem of areas that a three-dimensional simple-harmonic motion will lie in a plane.

Problem 32. Show from the theorem of areas that a plane simple-harmonic motion has a constant areal velocity, and find this areal velocity in terms of the initial conditions.

Problem 33. Show that the bob of a simple pendulum moves with simple-harmonic motion when the amplitude is small.

8. Motion under an Inverse-square Force.—One of the first uses to which Newton put his formulation of the laws of motion was the description of the motions of the planets. He found that the observed motions could be described by the statement that every planet attracts and is attracted by every other planet and the sun with a force which is proportional to the product of the masses of the two bodies and inversely proportional to the square of the distance between them. The planets are small enough, compared with the distances between them, so that they can be considered as particles for a first approximation. For more accurate results it will be shown later that the distances to be taken are those between centers.

The formal statement of this law of gravitation is that between every pair of particles there is a force of attraction whose magnitude is given by

$$F = G\frac{m_1 m_2}{r^2} \tag{2-26}$$

It will be shown later that the same law holds between larger spherical bodies whose density is a function of the radius only, if the distance r is measured between the centers.

a. Separation of the Center of Mass.—To write the differential equations of motion let x_1, y_1, z_1, and m_1 be the coordinates and mass of one body and x_2, y_2, z_2, and m_2 those of the other. The equations for the x coordinates are then

$$\left.\begin{aligned} m_1\frac{d^2x_1}{dt^2} &= \frac{-Gm_1m_2(x_1 - x_2)}{[(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2]^{\frac{3}{2}}} \\ m_2\frac{d^2x_2}{dt^2} &= \frac{-Gm_1m_2(x_2 - x_1)}{[(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2]^{\frac{3}{2}}} \end{aligned}\right\} \tag{2-27}$$

Now define the new coordinates

$$\left.\begin{aligned} \xi &= \frac{m_1x_1 + m_2x_2}{m_1 + m_2} \quad \eta = \frac{m_1y_1 + m_2y_2}{m_1 + m_2} \quad \zeta = \frac{m_1z_1 + m_2z_2}{m_1 + m_2} \\ X &= x_1 - \xi = \frac{m_2}{m_1 + m_2}(x_1 - x_2) \\ Y &= y_1 - \eta = \frac{m_2}{m_1 + m_2}(y_1 - y_2) \\ Z &= z_1 - \zeta = \frac{m_2}{m_1 + m_2}(z_1 - z_2) \end{aligned}\right\} \quad (2\text{-}28)$$

Combining equations (2-27) and (2-28) leads to the equations of motion in the new coordinates,

$$\frac{d^2\xi}{dt^2} = 0 \quad (2\text{-}29a)$$

$$\frac{d^2X}{dt^2} = -\frac{Gm_2^3}{(m_1 + m_2)^2}\frac{X}{(X^2 + Y^2 + Z^2)^{\frac{3}{2}}} = -\frac{KX}{(X^2 + Y^2 + Z^2)^{\frac{3}{2}}} \quad (2\text{-}29b)$$

and similar equations in the other two pairs of coordinates. Equation (2-29*a*) describes the motion of the center of mass, and (2-29*b*) describes the motion of the particle m_1 about the center of mass. The three differential equations in X, Y, Z are just the same as would be written for the motion of a single particle about a fixed center of force, when the constant K is given the proper significance. The ratios of the forces along the three axes are equal to the ratios of the coordinates X, Y, Z, so that the force is one of attraction toward the center of mass.

b. Motion about the Center of Mass under a Force toward It.— Motion of this kind satisfies the law of the conservation of angular momentum, and therefore the orbit of the particles, in the X, Y, Z coordinates, will be confined to a plane. Let this be the X-Y plane. It is then convenient to use plane polar coordinates (r,θ) in this plane. Hence let $Z \equiv 0$, and

$$X = r\cos\theta \qquad Y = r\sin\theta \quad (2\text{-}30)$$

In these coordinates the force is directed along r toward the origin and is proportional to K/r^2.

The equations of motion (2-29*b*), when expressed in these plane-polar coordinates, have the form

$$\frac{d^2r}{dt^2} - r\left(\frac{d\theta}{dt}\right)^2 = -\frac{K}{r^2} \qquad 2\frac{dr}{dt}\frac{d\theta}{dt} + r\frac{d^2\theta}{dt^2} = 0 \quad (2\text{-}31)$$

These two equations show clearly the complications that are met in transforming to other than rectangular coordinate systems. The second derivative of a coordinate is not always equal to the acceleration in the direction in which the coordinate is increasing. The first of the equations shows that, even when the distance r is constant, a force along r is necessary to maintain the motion. Hence there can be an acceleration along r, even though r does not change at all. The second equation gives the force and acceleration perpendicular to r. This, again, is not simply the second derivative of θ but involves also the first derivative of θ as well as the first derivative of r. In using these equations in polar coordinates, it is very important to keep these facts in mind.

Problem 34. Work out in detail the transformations leading to equations (2-29*a*) and (2-29*b*).

Problem 35. Carry out the transformation to obtain equations (2-31) from (2-2) and (2-30).

Problem 36. Express the theorem of angular momentum in these polar coordinates.

Problem 37. Multiply the second of equations (2-31) by $r\,dt$, and integrate. The result is one of Kepler's laws of motion.

To find the orbit of a particle moving around a center, equations (2-31) could be integrated, and then the time could be eliminated between the solutions. A more convenient method, however, is to eliminate the time in the differential equations and so to obtain a differential equation of the orbit. This may be done by noting that $dr/dt = (dr/d\theta)(d\theta/dt)$ and that, from Prob. 36, $r^2\,d\theta/dt = C$. If, then, $u = 1/r$, the result is

$$\frac{d^2u}{d\theta^2} + u = \frac{K}{C^2} \qquad (2\text{-}32)$$

If the right-hand side of this equation were zero, it would be of the same form as (2-23) and so would have the same form of

solution. It can be reduced to this form by subtracting the constant on the right from the dependent variable and calling the result a new dependent variable. The solution is

$$u = \frac{K}{C^2} + A \cos (\theta + \delta) \tag{2-33}$$

Problem 38. Obtain equation (2-32) from (2-31).

Problem 39. A conic section may be defined as a curve such that the distance from any point on it to a fixed point, called the *focus*,

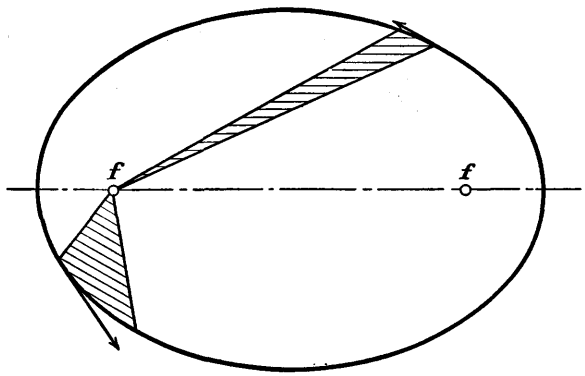


FIG. 2-4.—When the force of attraction is inversely proportional to the square of the distance from the attracting center, the orbit has this center at one focus. Note the difference between this case and that shown in Fig. 2-3.

divided by the distance from the same point on the curve to a fixed line, called the *directrix*, is equal to a constant ϵ, called the *eccentricity*. Using this definition, show that equation (2-33) represents a conic section with the focus at the origin (see Fig. 2-4). Show also that the eccentricity is equal to AC^2/K, in terms of the constants of equation (2-33).

Problem 40. If $\epsilon < 1$, the conic section is an ellipse. For this case show that the major and minor semiaxes are

$$a = \frac{KC^2}{K^2 - A^2C^4} \qquad b = \frac{C^2}{(K^2 - A^2C^4)^{\frac{1}{2}}} \tag{2-34}$$

Problem 41. Find the potential energy of a particle that is attracted toward the origin with a force inversely proportional to the square of the distance. The potential energy is set equal to zero for the particle at an infinite distance from the origin.

Problem 42. Find the total energy of a particle moving in an ellipse, and show that this energy can be expressed as a function of the major axis and the constants of the differential equations only and is independent of the minor axis.

Problem 43. Find the constants in equation (2-33) in terms of the initial position and velocity of the particle.

Problem 44. A particle is given an initial velocity perpendicular to the line connecting it with the attracting center. Find the values of this initial velocity for which the particle will move in a circle, for which it will fly off and never return, and for which it will move in an ellipse.

Problem 45. The area of an ellipse is πab, and the rate at which the area is covered is $C/2$. By using this fact, find the time necessary for a planet to go around the sun in terms of the constants describing the orbit, and show that the square of the period is proportional to the cube of the major axis.

CHAPTER III

LINEAR EQUATIONS OF ORDER HIGHER THAN THE FIRST

Most problems in physics require the solution of differential equations of higher than the first order. In the preceding chapter such equations were solved by reduction to first-order equations and subsequent integration in steps. This process is not always possible, and when it is possible it is not always the most convenient method. For the class of *linear* differential equations there is a general theory, so that in treating physical problems an attempt is usually made to approach the matter in such a way that the resulting differential equations are linear. Hence such equations are very important in mathematical physics, and this chapter will treat the methods necessary for the solution of some of the simpler linear equations.

1. General Properties of Linear Differential Equations.— The general linear differential equation is

$$\frac{d^n y}{dx^n} + P_1(x)\frac{d^{n-1}y}{dx^{n-1}} + P_2(x)\frac{d^{n-2}y}{dx^{n-2}} + \cdots P_n(x)y = Q(x) \quad (3\text{-}1)$$

Each term on the left-hand side contains y or a derivative of y to only the first power. The various coefficients, and the quantity Q that may appear on the right-hand side, are functions of x only. The equation as written above is a *nonhomogeneous equation.* When $Q(x)$ is omitted or set equal to zero, the equation is called a *homogeneous linear differential* equation. These two forms are of importance, for the general solution of the nonhomogeneous equations is obtained by adding a particular integral to the general solution of the corresponding homogeneous equation.

The property of linear equations that makes them so amenable to solution is contained in the following two statements:

1. The sum of any two solutions of a homogeneous linear

differential equation is also a solution of the equation, and the product of any solution by a constant is a solution.

2. If any solution of the homogeneous equation is added to a solution of the complete equation, the sum is another solution of the complete nonhomogeneous equation.

On the basis of these statements, any solution of a homogeneous linear differential equation can be multiplied by an arbitrary constant and still be a solution, and the general solution can be expressed as a sum of n terms, each of which is a solution multiplied by an arbitrary constant.

On the other hand, a solution of the nonhomogeneous equation cannot be multiplied by an arbitrary constant, for it would then no longer satisfy the equation. There can be added to it, however, a general solution of the corresponding homogeneous equation, which will contain n arbitrary constants. The sum will then be the general solution of the nonhomogeneous equation.

Consider, for example, the nonhomogeneous linear equation

$$x^2 \frac{d^2y}{dx^2} - x \frac{dy}{dx} + y = 2 \log x \tag{3-2}$$

The corresponding homogeneous equation is

$$x^2 \frac{d^2y}{dx^2} - x \frac{dy}{dx} + y = 0 \tag{3-2a}$$

One solution of this homogeneous equation is

$$y_1 = x \tag{3-2b}$$

as can be checked by substitution. This is still a solution when multiplied by any arbitrary constant.

Another solution of equation (3-2a) is

$$y_2 = x \log x \tag{3-2c}$$

so that a general solution of (3-2a) is

$$y = A_1 y_1 + A_2 y_2 = A_1 x + A_2 x \log x \tag{3-2d}$$

Each term of equation (3-2a) operates separately on the two terms of (3-2d) so that each solution separately gives zero

when substituted. The two terms in (3-2*d*) do not interfere with each other in any way so that the sum satisfies the differential equation. That this sum is a general solution follows from the presence of the two independent arbitrary constants.

To find the general solution of the complete equation one must find some particular integral of it. In the case of equation (3-2) this can be taken as

$$y_p = 2 \log x + 4 \tag{3-2e}$$

That this satisfies the differential equation can be shown by substitution, but it will not satisfy the equation when multiplied by an arbitrary constant. If the above y_p were multiplied by 3 and substituted in the left side of equation (3-2), the result would be three times the right side.

To get the general solution of the complete equation, it is necessary to add to the *particular integral* a *complementary function,* which is a general solution of the homogeneous equation. Since this complementary function makes the left side equal to zero, the complete equation is still satisfied.

Problem 1. Consider the equation

$$4x^2 \frac{d^2y}{dx^2} + y = 3x^2$$

Show that $y_1 = x^{\frac{1}{2}}$ and $y_2 = x^{\frac{1}{2}} \log x$ are solutions of the homogeneous equation and that $y_p = \frac{1}{3}x^2$ is a solution of the whole equation. Write the general solution, and find the solution that passes through ($x = 1$, $y = 1$) with zero slope.

In dealing with linear differential equations, it is often convenient to use the symbolic operator D. This represents the operation of taking a derivative, and its square represents the repeated operation, or the taking of the second derivative. In this notation

$$\frac{dy}{dx} \equiv Dy \qquad \frac{d^2y}{dx^2} \equiv D^2y \qquad \frac{dy}{dx} - ay \equiv (D - a)y$$

With this notation the general equation (3-1) becomes

$$D^n y + P_1(x)D^{n-1}y + P_2(x)D^{n-2}y + \cdots + P_n(x)y = Q(x) \tag{3-3}$$

For many purposes the operator D can be treated as an algebraic quantity and handled as such. In so doing special attention must be given to the fact that it does not commute with functions of x so that

$$f(x)Dy \neq Df(x)y = f(x)\frac{dy}{dx} + \frac{df(x)}{dx}y$$

2. Linear Differential Equations with Constant Coefficients. When the coefficients in the linear differential equation are constants rather than functions of x, the operator D can be treated quite generally as an algebraic quantity since it does commute with constants. The general equation may then be written

$$(D^n + a_1D^{n-1} + a_2D^{n-2} + \cdots a_n)y = f(D)y = Q \quad (3\text{-}4)$$

If the polynomial in D can be factored, the equation becomes

$$(D - \alpha_1)(D - \alpha_2)(D - \alpha_3) \cdots (D - \alpha_n)y = Q \quad (3\text{-}5)$$

where α_1, α_2, α_3, . . . , α_n are the n roots of the algebraic equation

$$D^n + a_1D^{n-1} + a_2D^{n-2} + \cdots a_n = 0 \quad (3\text{-}6)$$

in which D is treated as an algebraic number rather than an operator. Equation (3-6) is called the *auxiliary equation*, and a definite general procedure can be given for writing the solution of equation (3-4) after the roots of equation (3-6) have been determined.

The validity of the factoring indicated in equation (3-5) can be checked by substituting derivatives for the D's and carrying out the necessary operations. When only constant coefficients are present, the operator D can be handled with respect to addition, subtraction, and multiplication as a simple algebraic quantity. Division by D, however, requires further consideration.

3. Solution of the Homogeneous Equation with Constant Coefficients.—When the right-hand side of equation (3-5) is zero, the equation is homogeneous, and its general solution is the complementary function of the solution of the complete

equation. It can be seen immediately that if y_n is a solution of the equation

$$(D - \alpha_n)y_n = 0 \tag{3-7}$$

it is also a solution of

$$(D - \alpha_1)(D - \alpha_2)(D - \alpha_3) \cdots (D - \alpha_n)y = 0 \tag{3-8}$$

This follows because the operation with $(D - \alpha_n)$ gives zero, and the remaining operations cannot change this. Furthermore, the order of the factors in equation (3-8) is entirely arbitrary and can be changed at will, so that the solution of any one of the first-order linear equations

$$(D - \alpha_1)y = 0, \qquad (D - \alpha_2)y = 0, \ldots, (D - \alpha_n)y = 0,$$

is a solution of equation (3-8). These solutions are

$$e^{\alpha_1 x}, e^{\alpha_2 x}, e^{\alpha_3 x}, \ldots, e^{\alpha_n x}$$

and the general solution of the homogeneous differential equation is

$$y = A_1 e^{\alpha_1 x} + A_2 e^{\alpha_2 x} + A_3 e^{\alpha_3 x} \cdots A_n e^{\alpha_n x} \tag{3-9}$$

where the A's are arbitrary constants.

The procedure for finding the general solution of a homogeneous linear differential equation is first to solve the auxiliary equation and then to use the roots to form an expression of the form (3-9).

Consider, for example, the second-order equation

$$\frac{d^2y}{dx^2} + (a + b)\frac{dy}{dx} + aby = 0 \tag{3-10}$$

The auxiliary equation is

$$D^2 + (a + b)D + ab = 0 \tag{3-10a}$$

of which the factors are $(D + a)$ and $(D + b)$. Then the general solution is

$$y = A_1 e^{-ax} + A_2 e^{-bx} \tag{3-10b}$$

That this is a solution can be shown by substitution in the differential equation, and that it is the general solution follows from the presence of the two independent arbitrary constants.

A special case arises when one or more of the roots of the auxiliary equation are repeated. In this case there are only $n - 1$, or fewer, independent solutions of the simple exponential form, and another form must be found to complete the general solution. This can be formed by multiplying the exponential containing the repeated root by x. If the root is triple, a third solution can be formed by multiplying by x^2. Thus the general solution of

$$(D - \alpha)^2 y = 0 \tag{3-11}$$

is

$$y = A_1 e^{\alpha x} + A_2 x e^{\alpha x} \tag{3-11a}$$

and the general solution of

$$(D - \alpha)^3 y = 0 \tag{3-12}$$

is

$$y = A_1 e^{\alpha x} + A_2 x e^{\alpha x} + A_3 x^2 e^{\alpha x} \tag{3-12a}$$

Problem 2. Find the solution of

$$\frac{d^2y}{dx^2} = m^2 y$$

that passes through the origin with unit slope.

Problem 3. Write the general solution of

$$\frac{d^3y}{dx^3} - 3\frac{d^2y}{dx^2} + 4y = 0$$

Problem 4. Write the general solution of

$$\frac{d^2y}{dx^2} + 8\frac{dy}{dx} + 16y = 0$$

and evaluate the arbitrary constants to give a solution that will pass through the origin and through the point ($x = 2$, $y = 2$).

4. Solution of the Nonhomogeneous Equation with Constant Coefficients.—As was pointed out above, the solution of the nonhomogeneous equation is the sum of a *complementary function* and a *particular integral*. The complementary function is a general solution of the corresponding homogeneous equation and can be found by the method just indicated. The particular integral must be found by some other means. In some cases one can be seen by inspection. For several special forms of the

function Q on the right-hand side there are special methods that can be used, and a general procedure can be given for use when the special methods do not apply. This section will be devoted to these methods of finding a particular integral.

a. $$Q = C_0 + C_1x + \cdots + C_Nx^N$$

When the right-hand side of the equation is a finite polynomial in positive powers of x, the particular integral can also be a finite polynomial in positive powers of x. In most cases the highest power of x in the solution will not be greater than N, the highest power in the equation. If, however, the operator D appears as a factor in the auxiliary equation, the highest power of x in the solution will be greater than N by the number of times D appears as a factor.

For example, consider the equation

$$\frac{d^3y}{dx^3} + p\frac{d^2y}{dx^2} + q\frac{dy}{dx} + ry = C_0 + C_1x + C_2x^2 \qquad (3\text{-}13)$$

Assume as a particular integral

$$y = b_0 + b_1x + b_2x^2 \qquad (3\text{-}13a)$$

Substituting in the differential equation and equating the coefficients of each power of x leads to

$$\begin{aligned} rb_2 &= C_2 \\ rb_1 + 2qb_2 &= C_1 \\ rb_0 + qb_1 + 2pb_2 &= C_0 \end{aligned}$$

Hence

$$b_2 = \frac{C_2}{r} \qquad b_1 = \frac{C_1r - 2qC_2}{r^2}$$

$$b_0 = \frac{C_0r^2 - 2pC_2r - qC_1r + 2q^2C_2}{r^3}$$

With these values of the coefficients the expression (3-13*a*) is a particular integral of equation (3-13). There is no need to investigate whether there may be other solutions, for when this particular integral is added to the complementary function the sum is the general solution of the differential equation.

As an example of the case in which D appears as a factor in the auxiliary equation consider the equation

$$\frac{d^2y}{dx^2} - a\frac{dy}{dx} = x \tag{3-14}$$

A particular integral of this is

$$y_p = -\frac{x}{a^2} - \frac{x^2}{2a} \tag{3-14a}$$

This contains a term in x^2 that is one degree higher than the term in x on the right-hand side of equation (3-14).

b. $$Q = Ae^{ax}$$

When the right-hand side of the equation is an exponential such that the coefficient of the exponent is not a root of the auxiliary equation, the solution can be written down at once.

If $f(D)y = Ae^{ax}$

$$y_p = \frac{Ae^{ax}}{f(a)} \tag{3-15}$$

where the denominator is merely $f(D)$ with a substituted for D. If, however, a is a root of the auxiliary equation, $f(a) = 0$ and the form indicated is not useful. In this case the particular integral is xe^{ax} multiplied by a suitable coefficient that can be determined by substitution in the differential equation.

As an example of this special case, consider the equation

$$\frac{d^2y}{dx^2} - a^2y = e^{ax} \tag{3-16}$$

Assume the particular integral in the form

$$y_p = Cxe^{ax} \tag{3-16a}$$

Substitution in the differential equation leads to

$$C = \frac{1}{2a}$$

so that the general solution is

$$y = A_1e^{ax} + A_2e^{-ax} + \frac{x}{2a}e^{ax} \tag{3-16b}$$

c. $Q = c \sin ax$ or $Q = c \cos ax$

If the right-hand side of the equation contains sine or cosine terms that are not solutions of the homogeneous equation, a particular integral can be built up of these quantities with coefficients to be determined by substitution in the differential equations. If the trigonometric functions are solutions of the homogeneous equation, they must be multiplied by x before substitution to determine the coefficients.

For example, take the equation

$$\frac{d^2y}{dx^2} + 2\frac{dy}{dx} + y = \sin ax \tag{3-17}$$

The complementary function is

$$y_c = (A_1 + A_2x)e^{-x} \tag{3-17a}$$

so that sin ax is not a solution of the homogeneous equation. To find a particular integral let

$$y_p = c_1 \sin ax + c_2 \cos ax \tag{3-17b}$$

Substitution in the differential equation leads to

$$[(1 - a^2)c_1 - 2ac_2] \sin ax + [2ac_1 + (1 - a^2)c_2] \cos ax = \sin ax$$

Since this equation must hold for all values of x, the coefficients of the terms in sin ax and cos ax must separately be equal. This condition leads to

$$\begin{aligned}(1 - a^2)c_1 - 2ac_2 &= 1\\ 2ac_1 + (1 - a^2)c_2 &= 0\end{aligned}$$

From these the particular integral is

$$y_p = \frac{1}{(1 + a^2)^2}[(1 - a^2) \sin ax - 2a \cos ax] \tag{3-17c}$$

As an example of the case in which the sine and cosine terms are solutions of the homogeneous equation, take the simple equation

$$\frac{d^2y}{dx^2} + a^2y = \sin ax \tag{3-18}$$

Let the particular integral have the form

$$y_p = c_1 x \sin ax + c_2 x \cos ax \tag{3-18a}$$

Substitution in the differential equation shows that c_1 must be zero and that a particular integral is

$$y_p = -\frac{x}{2a} \cos ax \tag{3-18b}$$

d. The General Case.—Although special methods can be written down for finding particular integrals of many of the commonly appearing equations, it is convenient to have available a general method which, even though it may be a little complicated, can be applied to any case that may arise. Such a method can be formulated in terms of the operator D.

Consider the equation

$$f(D)y = (D - \alpha_1)(D - \alpha_2) \cdots (D - \alpha_n)y = Q \tag{3-19}$$

If the operator D is treated as an algebraic quantity, one may write

$$y = \frac{1}{(D - \alpha_1)(D - \alpha_2) \cdots (D - \alpha_n)} Q \tag{3-20}$$

This is now a symbolic form of a particular integral, but its significance must be interpreted. One way of treating this form s to continue to regard D as an algebraic quantity and to separate the expression (3-20) into partial fractions. This leads to

$$y = \frac{N_1}{D - \alpha_1} Q + \frac{N_2}{D - \alpha_2} Q + \cdots \frac{N_n}{D - \alpha_n} Q \tag{3-21}$$

In equation (3-21) the quantities $N_1, N_2, \cdots, N_n$ are coefficients such that the right sides of (3-20) and (3-21) are equal. If, for example,

$$y = \frac{1}{D^2 - m^2} Q = \frac{N_1}{D + m} Q + \frac{N_2}{D - m} Q \tag{3-22}$$

the values of N_1 and N_2 can be determined by recombining the partial fractions. This leads to

$$y = \frac{(N_1 + N_2)D - (N_1 - N_2)m}{D^2 - m^2} Q \tag{3-23}$$

which must be the same as the first equation (3-22). Since D is an operator, it cannot be canceled by a number so that there are two equations for the determination of N_1 and N_2.

The solution of these is

$$N_1 = -\frac{1}{2m} \qquad N_2 = \frac{1}{2m}$$

It is now necessary to interpret an expression of the form

$$y = \frac{C}{D - \alpha_n} Q \tag{3-24}$$

which is just the symbolic form of the solution of a first-order linear equation. This was treated in Chap. I. Hence

$$\frac{C}{D - \alpha_n} Q(x) = Ce^{\alpha_n x} \int^x Q(u) e^{-\alpha_n u}\, du \tag{3-25}$$

With this interpretation, equation (3-22) can be immediately evaluated in terms of integrals.

To illustrate the procedure take the second-order equation

$$\frac{d^2y}{dx^2} - 7\frac{dy}{dx} + 12y = ax \tag{3-26}$$

Of course, this could be handled by the special method described above for this form of equation, but it will also serve as an illustration of the general method. The particular integral can be symbolically written as

$$y = \frac{ax}{(D - 3)(D - 4)} = \frac{ax}{D - 4} - \frac{ax}{D - 3} \tag{3-26a}$$

This has now been reduced to a sum of two quantities of the form (3-25) so that

$$\begin{aligned} y &= ae^{4x} \int^x ue^{-4u}\, du - ae^{3x} \int^x ue^{-3u}\, du \\ &= -\frac{ax}{4} - \frac{a}{16} + \frac{ax}{3} + \frac{a}{9} = \frac{ax}{12} + \frac{7a}{144} \end{aligned} \tag{3-26b}$$

If two or more of the roots are repeated, the separation into partial fractions cannot be carried out completely and there will remain expressions of the form

$$y = \frac{1}{(D - \alpha_n)^2} Q$$

This can be interpreted by successive integration

$$y = \frac{1}{D - \alpha_n}\left(\frac{Q}{D - \alpha_n}\right) = \frac{e^{\alpha_n x}}{D - \alpha_n}\int^x Q(u)e^{-\alpha_n u}\,du$$
$$= e^{\alpha_n x}\int^x e^{-\alpha_n v}\left(e^{\alpha_n v}\int^v Q(u)e^{-\alpha_n u}\,du\right)dv$$
$$= e^{\alpha_n x}\int^x dv\int^v Q(u)e^{-\alpha_n u}\,du \qquad (3\text{-}27a)$$

e. The Green's Function.—It is also possible to approach the particular integral in a somewhat different way, which is often instructive in physical problems. For simplicity in exposition, consider a second-order equation only. The particular integral as shown in equations (3-21) and (3-25) has the form

$$y_p = N_1 e^{\alpha_1 x}\int^x Q(u)e^{-\alpha_1 u}\,du + N_2 e^{\alpha_2 x}\int^x Q(u)e^{-\alpha_2 u}\,du \qquad (3\text{-}28a)$$

This can be combined into a single integral

$$y_p = \int^x Q(u)\,[N_1 e^{\alpha_1(x-u)} + N_2 e^{\alpha_2(x-u)}]du$$
$$= \int^x Q(u)Y(x,u)du \qquad (3\text{-}28b)$$

The function $Y(x,u)$ is known as the *Green's function* for the differential equation. It has the following properties:

1. Considered as a function of x, $Y(x,u)$ is a solution of the homogeneous differential equation for any value of the parameter u.

2. $Y(x,u) = 0$ when $x = u$

This is true in equation (3-28*b*) because $N_1 + N_2 = 0$.

3. $\partial Y/\partial x = 1$ when $x = u$

This is true in equation (3-28*b*) because

$$N_1\alpha_1 + N_2\alpha_2 = 1$$

which follows from the separation into partial fractions.

If conditions 2 and 3 are imposed on the arbitrary constants of a general solution of the homogeneous equation, the result is a Green's function suitable for use in equation (3-28*b*). For an example, consider the equation

$$\frac{d^2y}{dx^2} + 6\frac{dy}{dx} + 8y = x^2 \qquad (3\text{-}29)$$

The general solution of the homogeneous equation is

$$y = Ae^{-2x} + Be^{-4x} \tag{3-29a}$$

and conditions 2 and 3 give

$$\begin{aligned} Ae^{-2u} + Be^{-4u} &= 0 \\ -2Ae^{-2u} - 4Be^{-4u} &= 1 \end{aligned}$$

Solving for A and B and inserting in (3-29a) give

$$Y(x,u) = \tfrac{1}{2}e^{-2(x-u)} - \tfrac{1}{2}e^{-4(x-u)} \tag{3-29b}$$

The particular integral is then

$$\begin{aligned} y_p &= \tfrac{1}{2}\int^x u^2[e^{-2(x-u)} - e^{-4(x-u)}]du \\ &= \tfrac{1}{8}x^2 - \tfrac{3}{16}x + \tfrac{7}{64} \end{aligned} \tag{3-29c}$$

The Green's function representation of the particular integral is often instructive in mechanical problems. Consider the equation for a forced oscillator,

$$\frac{d^2x}{dt^2} + p\frac{dx}{dt} + rx = Q(t) \tag{3-30}$$

The quantity on the right-hand side, $Q(t)$, is $1/m$ times an external force in addition to the forces of the system represented by $-mp\ dx/dt$ and $-mrx$. If it is assumed that the external force acts only at the time τ and for the short space of time $d\tau$, it will produce a velocity $Q(\tau)d\tau$. A solution of the equation can be written such that $x = 0$ at $t = \tau$ and $\dot{x} = Q(\tau)d\tau$ at $t = \tau$. This solution is then $Q(\tau)Y(t,\tau)d\tau$; and if the force continues to act, the motion due to the force at other times is merely

$$x_p = \int^t Q(\tau)Y(t,\tau)d\tau \tag{3-30a}$$

In all this work the lower limit of the integral is omitted. The inclusion of a lower limit would be equivalent to the addition of terms of the same form as those in the complementary function.

Problem 5. Find the general solution of

$$\frac{d^2y}{dx^2} + 5\frac{dy}{dx} + 6y = e^{4x}$$

and evaluate the constants to find the solution that passes through the origin with unit slope.

Problem 6. Write the general solution of

$$\frac{d^2y}{dx^2} - y = 2 + 5x$$

Problem 7. Write the general solution of

$$\frac{d^2y}{dx^2} + 2\frac{dy}{dx} + y = 2e^{2x}$$

Problem 8. Find the general solution of

$$\frac{d^2y}{dx^2} + 2\frac{dy}{dx} + y = \frac{1 + x^2}{(1 + x)^3}$$

Problem 9. Show that one gets the general solution correctly without considering the complementary function, if the constants of integration are included in the various integrals of the form (3-25) and (3-27).

Problem 10. Find the solution of

$$\frac{d^3y}{dx^3} + 4\frac{d^2y}{dx^2} + 4\frac{dy}{dx} = 5$$

that passes through the origin.

5. Complex Numbers.—Many equations of mathematical physics, especially those representing vibrating systems, can be most easily solved by the use of complex numbers. In this section and the next two, we shall deal with some of the more elementary properties of these numbers.

In elementary mathematics, one deals first with positive real integers only. With these it is always possible to carry out the operation of addition. However, in order that it shall always be possible to carry out the operation of subtraction, it is necessary to introduce the negative real integers. These numbers are not sufficient to permit the operation of division always to be carried out, and all the positive and negative rational fractions must also be included. This system of positive and negative rational numbers will always permit the fundamental operations of addition and subtraction, multiplication, and division to be carried out, provided that the division

by zero is excluded or specially defined. However, there are still other operations that cannot be carried out in this number system. One of these is the extraction of roots, even of positive numbers. To permit this and other "irrational" operations, it is necessary to introduce the irrational numbers. Then one has all the real numbers between positive and negative infinity, and they can be correlated with the points on a line. Still, however, the system is incomplete, since it does not permit the extraction of the roots of the negative numbers. There is no number that is equal to the square root of -1. It is necessary, therefore, to define such a number, which is called i.

$$i = +\sqrt{-1}$$

The other square root of -1 is $-i$. This i is called the *unit of imaginary numbers*, although it is no more imaginary, in the common sense of the word, than negative or irrational numbers. Any imaginary number is a product of a real number and the imaginary unit, such as $3i$ or ai for instance.

Two imaginary numbers can be added or subtracted by adding or subtracting the real coefficients and multiplying the results by i. To multiply or divide two imaginary numbers, it is necessary to multiply or divide both the coefficients and the units.

A complex number is the sum of a real and an imaginary number, such as $a + ib$. It is customary to call a the *real part* and b (not ib) the *imaginary part* of this number. Two complex numbers are equal when their real parts and their imaginary parts are equal, respectively. In addition or subtraction, the real parts and the imaginary parts are added or subtracted separately. In multiplication, the complex numbers are treated as binomials.

A complex number z_1 is said to be the *complex conjugate* of another complex number z_2 if the real part of z_1 is equal to the real part of z_2 and the imaginary part of z_1 is the negative of the imaginary part of z_2.

Since the real numbers and the imaginary numbers are special cases of complex numbers, this definition of complex

numbers has extended very materially the variety of numbers available for calculations. In this set of complex numbers every algebraic equation has a number of roots equal to its degree, and the range of solutions of differential equations is very materially extended.

On the basis of the above definitions, a number of properties of complex numbers follow directly. Among these are the following:

1. The product of an even number of imaginary numbers is real, and the product of an odd number of imaginary numbers is imaginary.

2. The product of a complex number and its complex conjugate is real.

3. The sum of a number and its complex conjugate is real, and the difference is imaginary. The sum is twice the real part of the complex number, and the difference is $2i$ times the imaginary part.

It is often convenient to represent real numbers by points on a line. In a similar way, it is often convenient to represent complex numbers by points on a plane. If the points are designated by their rectangular coordinates, the point (x,y) represents the complex number $(x + iy)$.

The distance from the origin to the point (x,y) is called the *modulus* or the *absolute value* of the number. The angle between the straight line from the origin to the point and the axis of real numbers is called the *argument* of the number.

Problem 11. Show that $(x + iy) = r(\cos \varphi + i \sin \varphi)$, where r is the modulus and φ is the argument of the complex number $(x + iy)$.

Problem 12. Show that

$$z_1 z_2 = r_1 r_2[\cos (\varphi_1 + \varphi_2) + i \sin (\varphi_1 + \varphi_2)]$$

where r_1, r_2 and φ_1, φ_2 are the moduli and the arguments of z_1 and z_2, respectively.

Problem 13. Show that

$$\frac{a + ib}{c + id} = \frac{ac + bd}{c^2 + d^2} + i \frac{bc - ad}{c^2 + d^2}$$

in the sense that the right-hand side multiplied by the denominator of the left-hand side is equal to the numerator of the left-hand side.

Problem 14. Show that

$$\frac{z_1}{z_2} = \frac{r_1}{r_2}[\cos(\varphi_1 - \varphi_2) + i \sin(\varphi_1 - \varphi_2)]$$

in the sense of the previous problem.

Problem 15. Show that

$$z^{1/n} = r^{1/n}\left(\cos\frac{\varphi + 2\pi m}{n} + i \sin\frac{\varphi + 2\pi m}{n}\right)$$

in the sense that the nth power of the right-hand side is equal to z. In this expression, n and m are integers. The n different roots are obtained by using the n different values of m.

6. Complex Functions of Real Variables.—As with real variables, a function of a complex variable is most generally defined as a table of values such that with every value of one complex number called the *argument* there is associated another complex number called the *function*. In some special cases the function can be defined by specifying a series of operations to be performed upon the argument and defining the result of these operations to be the function.

Although very general concepts of a function can be defined and are sometimes useful, it is necessary in dealing with complex numbers, just as in dealing with real numbers, to restrict considerably the nature of the functions used, if elementary methods of analysis are to be applied. Among the functions that are of importance are complex functions of a single real variable. These can be put into the form

$$z = f(x) = f_1(x) + if_2(x) \tag{3-31}$$

where x is the real variable. The usual criteria of continuity and differentiability can then be applied by applying them separately to the two functions $f_1(x)$ and $f_2(x)$.

Problem 16. If the function e^z of the complex variable z is defined by the infinite series

$$e^z = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \cdots$$

show that

$$e^{ix} = \cos x + i \sin x \tag{3-32}$$

when x is real. As a result of this fact a complex number can be written $z = re^{i\phi}$, where r is the modulus and φ is the argument of z.

Problem 17. Show that

$$\sin x = \frac{e^{ix} - e^{-ix}}{2i} \qquad \text{and} \qquad \cos x = \frac{e^{ix} + e^{-ix}}{2}$$

7. Hyperbolic Functions.—Many differential equations in physics, particularly those pertaining to electrical or mechanical problems, have as solutions functions known as *hyperbolic functions.* Without going into the many interesting properties of these functions, the definitions and some of the more elementary properties will be stated here. The functions can be defined as the solutions of the differential equations that they satisfy; but since they can also be expressed in terms of exponential functions, it is more convenient to define them in these terms. Hence, we have the following definitions:

$$\text{The hyperbolic sine of } x = \sinh x = -\sinh(-x) = \frac{e^{x} - e^{-x}}{2}$$

$$\text{The hyperbolic cosine of } x = \cosh x = \cosh(-x) = \frac{e^{x} + e^{-x}}{2}$$

$$\text{The hyperbolic tangent of } x = \tanh x = \frac{\sinh x}{\cosh x}$$

The three functions that are the reciprocals of these are correspondingly called the *hyperbolic cosecant,* the *hyperbolic secant,* and the *hyperbolic cotangent.* The above definitions combined with Prob. 17 show the relationship between the hyperbolic functions and the circular trigonometric functions. These hyperbolic functions have many properties that are similar to those of the circular functions; and although these properties can in all cases be derived from the above definitions, it is often convenient to remember them as properties of the hyperbolic functions. The following problems give some of these properties:

Problem 18. By determining the values at the points 0, 1, and infinity, sketch roughly the curves of the hyperbolic functions.

Problem 19. Show that $d(\sinh x)/dx = \cosh x$.

Problem 20. Show that $d(\cosh x)/dx = \sinh x$.

Problem 21. Show that $A \cosh x + B \sinh x$ is the general solution of the equation $d^2y/dx^2 - y = 0$.

Problem 22. Show that $\cosh^2 x - \sinh^2 x = 1$.

Problem 23. Show that $\sinh (ax + bx) = \sinh ax \cosh bx + \cosh ax \sinh bx$.

Problem 24. Show that $\sin ix = i \sinh x$; $\cos ix = \cosh x$.

8. Principles of Superposition and Decomposition.—As has already been indicated, any number of solutions of a homogeneous linear differential equation can be added together, or *superposed*, to form another solution. If the solutions are linearly independent, their addition after multiplication by arbitrary constants leads to the general solution.

In the case of nonhomogeneous linear differential equations, a principle of superposition can be stated as follows:

If

$$f(D)y_1 = Q_1 \qquad \text{and} \qquad f(D)y_2 = Q_2$$

then

$$f(D)(y_1 + y_2) = Q_1 + Q_2 \tag{3-33}$$

This shows that it is possible to separate the right-hand side of an equation into parts and, after finding an integral corresponding to each part, to add them together to get an integral of the original equation.

For example, if the equation under consideration is

$$(D^2 + a^2)y = e^{mx} + C \tag{3-34}$$

each part of the right-hand side may be treated separately. $y_1 = e^{mx}/(m^2 + a^2)$ is a particular integral when only the term e^{mx} is considered, and $y_2 = C/a^2$ is a particular integral when only the constant C is considered. The sum of these is then a particular integral of the whole equation

$$y_p = y_1 + y_2 = \frac{e^{mx}}{m^2 + a^2} + \frac{C}{a^2} \tag{3-34a}$$

This principle of superposition is of very extensive use in physical problems, since it is the mathematical representation of the fact that many physical phenomena can be treated as superpositions of independent phenomena that do not interfere with

one another. For example, if two light waves pass through the same portion of space, the phenomenon is a superposition of the two waves without either influencing the other. In a similar way the terms that make some equations of physics nonhomogeneous are due to *sources*, such as sources of radiation. The principle of superposition shows that, as long as the differential equations are linear, the sources act entirely independently.

The principle of *decomposition* is somewhat more limited in its applicability than the principle of superposition. In general, it is not possible to know in advance how a solution will divide into partial solutions corresponding to the different parts of the nonhomogeneous term. In the special case, however, that the coefficients of the differential equation are all real and the term on the right side is complex, the particular integral will be complex. It then follows that the real part of the integral corresponds to the real part of the right-hand side and the imaginary part corresponds to the imaginary part of the right-hand side. As an example, consider the equation

$$\frac{d^2y}{dx^2} + a^2y = e^{ix} \tag{3-35}$$

The particular integral for the exponential form is then

$$y_p = \frac{e^{ix}}{i^2 + a^2} = \frac{e^{ix}}{a^2 - 1} \tag{3-35a}$$

Suppose now it is desired to find the solution of the equation

$$\frac{d^2y}{dx^2} + a^2y = \cos x \tag{3-36}$$

The right-hand side of this is just the real part of the right-hand side of equation (3-35); hence its particular integral is the real part of (3-35*a*).

$$y_p = \frac{\cos x}{a^2 - 1} \tag{3-36a}$$

The general solution is then

$$y = A \cos x + B \sin x + \frac{\cos x}{a^2 - 1} \tag{3-36b}$$

Furthermore, the solution of

$$\frac{d^2y}{dx^2} + a^2y = \sin x \tag{3-37}$$

is

$$y = A \cos ax + B \sin ax + \frac{\sin x}{a^2 - 1} \tag{3-37a}$$

where A and B are arbitrary constants.

When presented with an equation such as (3-36) or (3-37), it is often convenient to take another equation such as (3-35) that contains the equation in question as its real or its imaginary part and that may be easier to solve.

Problem 25. Find the solution of

$$\frac{d^2y}{dx^2} - 2\frac{dy}{dx} + y = x^2e^{3x}$$

that is equal to zero when $x = 0$ and when $x = 10$.

Problem 26. Use the principle of decomposition to find an integral of the equation

$$\frac{d^2y}{dx^2} - 2\frac{dy}{dx} + 4y = e^x \cos x$$

Problem 27. Use the principle of decomposition to find an integral of the equation

$$\frac{d^2y}{dx^2} + 2\frac{dy}{dx} + 4y = 3 \sin 2x$$

Problem 28. Find the general solution of

$$\frac{d^2y}{dx^2} + 2\frac{dy}{dx} + 4y = e^{2x}$$

Problem 29. Find the general solution of

$$\frac{d^2y}{dx^2} - a^2y = e^{ax} + e^{nx}$$

Problem 30. Find the general solution of

$$\frac{d^2y}{dx^2} - y = (e^x + 1)^2$$

Problem 31. Find an integral of

$$\frac{d^2y}{dx^2} - 2\frac{dy}{dx} - 3y = \cos^2 x$$

9. Equations with Variable Coefficients.—The solution of linear equations of order higher than the first, in which the coefficients are not constant, is required in the treatment of many physical problems. Many such equations have been studied and the properties of their solutions determined. The general discussion of such equations is quite elaborate, but a few suggestions will be given here to indicate how information can be obtained regarding the nature of the solutions in the neighborhood of certain values of the independent variable.

The general equation of this type has the form

$$p_0\frac{d^ny}{dx^n} + p_1\frac{d^{n-1}y}{dx^{n-1}} + p_2\frac{d^{n-2}y}{dx^{n-2}} + \cdots p_ny = Q \quad (3\text{-}38)$$

where $p_0, p_1, p_2, \ldots, p_n$ are functions of x only. Just as in the case of the equations with constant coefficients, the general solution of this equation is the sum of a particular integral and a complementary function, which is the general solution of the homogeneous equation. Methods for finding a particular integral will not be discussed here, and attention will be concentrated on the homogeneous equation.

The behavior of the solution is governed largely by the behavior of the coefficients of the equation for different values of x. In examining the behavior of the coefficients, the equation should be reduced to a standard form by dividing out $p_0(x)$. This gives

$$\frac{d^ny}{dx^n} + P_1\frac{d^{n-1}y}{dx^{n-1}} + P_2\frac{d^{n-2}y}{dx^{n-2}} + \cdots P_ny = 0 \quad (3\text{-}38a)$$

where $P_r = p_r/p_0$ and Q is taken to be equal to zero.

If all the $P_r(x)$'s are regular at the point $x = 0$, the general solution of (3-38a) is regular at this point and the point is called an *ordinary point* of the differential equation. A function $P(x)$

is said to be regular at $x = c$ if it can be expanded in a Taylor series,

$$P(x) = a_0 + a_1(x - c) + a_2(x - c)^2 + \cdots$$

about this point. Regularity requires, among other things, that a function be single-valued, continuous, and differentiable. A polynomial containing only positive powers is regular at all finite values of its argument. It may be regarded as a power series, most of whose coefficients are zero.

Points other than ordinary points are called *singular points* or *singularities*. If all the $p_r(x)$'s in equation (3-38) are regular, the only singularities in the finite region are at zeros of $p_0(x)$.

Consider the equation

$$(1 + x^2)\frac{d^2y}{dx^2} + 3x\frac{dy}{dx} + y = 0 \qquad (3\text{-}39)$$

All the coefficients are polynomials and regular for all finite values of x. When this is reduced to the standard form, it becomes

$$\frac{d^2y}{dx^2} + \frac{3x}{1 + x^2}\frac{dx}{dy} + \frac{1}{1 + x^2}y = 0 \qquad (3\text{-}39a)$$

In this form it can be seen that all the finite real values of x are ordinary points and that the only singularities in the finite region are at $x = \pm i$. The methods for finding solutions at ordinary points and at singularities will be considered separately.

10. Solutions in Power Series around Ordinary Points.—It is shown in treatises on differential equations that, in the neighborhood of an ordinary point $x = c$, the general solution of a linear homogeneous differential equation can be expressed as a power series with only positive integral powers.

$$y = \sum_{n=0}^{\infty} a_n(x - c)^n \qquad (3\text{-}40)$$

The values of the coefficients a_n can be determined by substitution in the differential equation itself. In practice, it is almost always most convenient to expand about the point $x = 0$. This can always be done by first making a linear change in

the independent variable. In further illustrations this will be assumed to have been done so that c can be set equal to zero in equation (3-40).

Another important practical consideration is that, while the standard form of equation (3-39a) is appropriate for determining the location and nature of the singularities, a form like that of equation (3-39), in which one has cleared of fractions and expressed each coefficient as a polynomial or power series in x, is the one to be used in finding the coefficients of the power-series solution.

To illustrate the procedure, let the solution of equation (3-39) be expressed in the form

$$y = \sum_{n=0}^{\infty} a_n x^n$$

When this is substituted in the differential equation, the result is

$$\sum_{n=0}^{\infty} [a_n n(n-1)x^{n-2} + a_n(n+1)^2 x^n] = 0 \qquad (3\text{-}39b)$$

In order that this power series be equal to zero for all values of x, it is both necessary and sufficient that the coefficient of each power of x be zero.

Consider the coefficient of x^{s-2}. In the first part of equation (3-39b), this appears when $n = s$ so that $s(s-1)a_s$ is a part of the desired coefficient. In the second part, x^{s-2} appears when $n = s - 2$. Hence a second part of the coefficient is $(s-1)^2 a_{s-2}$. The requirement is then that

$$s(s-1)a_s + (s-1)^2 a_{s-2} = 0 \qquad (3\text{-}39c)$$

for all values of s from 2 to infinity. Equation (3-39c) is the recursion formula for the coefficients in the power series.

The lowest value of s that can be used in the recursion formula is $s = 2$, for this gives the coefficient of x^0 in equation (3-39b). With $s = 2$ the formula gives a_2 in terms of a_0. With $s = 3$ it gives a_3 in terms of a_1. Other values of s give

a_4 in terms of a_2 and hence of a_0, and a_5 in terms of a_3 and hence of a_1. Proceeding in this way, all the coefficients can be determined in terms of a_0 and a_1. These two then become the arbitrary constants of the general solution, and the complete theory of linear differential equations shows that the series will converge in the neighborhood of the ordinary point. Hence the series constitutes a general solution.

Since a_0 and a_1 are arbitrary, $a_0 = 1$, $a_1 = 0$ gives a solution, which in this particular case contains only even powers of x. From the recursion formula it follows that $a_2 = -a_0/2 = -\frac{1}{2}$, and $a_4 = -9a_2/12 = \frac{3}{8}$. This solution is then

$$y_0 = 1 - \frac{x^2}{2} + \frac{3x^4}{8} - \cdots \tag{3-39d}$$

Similarly, $a_0 = 0$, $a_1 = 1$ gives another solution,

$$y_1 = x - \frac{2x^3}{3} + \frac{8x^5}{15} - \cdots \tag{3-39e}$$

The general solution can then be expressed as a linear combination of these two particular solutions.

$$y = a_0y_0 + a_1y_1 \tag{3-39f}$$

By arguments along the lines of the above illustration, it can be shown that a general solution of an nth order equation, in the neighborhood of an ordinary point, can be expressed in the form

$$y = a_0y_0 + a_1y_1 + a_2y_2 + \cdots + a_{n-1}y_{n-1} \tag{3-41}$$

where y_r is a power series beginning with x^r.

Problem 32. Work out a few more terms of the series solution of equation (3-39). Identify the function y_0 of equation (3-39*d*) as the series expansion of a binomial, and show by substitution that the binomial satisfies the differential equation.

Problem 33. Find the expansion in a power series around the origin of the solutions of Legendre's equation

$$(1 - x^2)\frac{d^2y}{dx^2} - 2x\frac{dy}{dx} + n(n + 1)y = 0 \tag{3-42}$$

Show that, for integral values of n, the series of even powers is a polynomial, with a finite number of terms, when n is even and that similarly the series of odd powers is finite when n is odd. These polynomials are the Legendre polynomials.

Problem 34. Find a solution of the equation

$$\frac{d^3y}{dx^3} + x^2\frac{d^2y}{dx^2} + ax\frac{dy}{dx} + by = 0 \tag{3-43}$$

Problem 35. Find the power-series solutions of

$$(1 - x^2)\frac{d^2y}{dx^2} + \frac{dy}{dx} + y = 0 \tag{3-44}$$

As an illustration of a situation that sometimes arises in evaluating the constants of a power-series solution, consider the equation

$$\frac{d^2y}{dx^2} + xy = 0 \tag{3-45}$$

Substitution of a power series leads to

$$\sum_{n=0}^{\infty} [n(n-1)a_n x^{n-2} + a_n x^{n+1}] = 0 \tag{3-45a}$$

A convenient form of recursion formula is obtained from this by picking out the coefficient of x^{s-2}. This leads to

$$s(s-1)a_s + a_{s-3} = 0 \tag{3-45b}$$

provided that s is equal to 3 or more. When $s = 3$, equation (3-45b) represents the coefficient of x and not the constant; if s is taken equal to 2, the recursion formula would seem to suggest a coefficient a_{-1}, which does not exist. If the coefficient of x^0, or the constant term, is picked out of (3-45a), it leads to $2a_2 = 0$. This might also have been obtained from (3-45b) by understanding that $a_{-1} = 0$. In any case of doubt, however, the situation can be clarified by referring to the sum of the type (3-45a).

When $s = 3$, equation (3-45b) gives $a_3 = -a_0/6$; when $s = 4$, $a_4 = -a_1/12$. But when $s = 5$, it leads to $a_5 = -a_2/20$ and a_2 must be zero. Hence not all powers of x appear in the

general solution, which, however, can still be written in the general form for a second-order equation,

$$y = a_0 y_0 + a_1 y_1 \tag{3-45c}$$

Problem 36. Write a general solution of

$$\frac{d^2y}{dx^2} + x^2 y = 0 \tag{3-46}$$

Problem 37. Find the first few terms of the power series for a general solution of

$$\frac{d^3y}{dx^3} + x\frac{dy}{dx} + bxy = 0 \tag{3-47}$$

11. Solutions in Power Series around Regular Singularities. There are some singularities in the neighborhood of which it is possible to write power-series solutions. These are called *regular points* or *regular singularities*. Other singular points, called *essential singularities*, will not be treated here.

The nature of a singularity can be determined by inspection of the coefficients of the differential equation when it is written in the standard form of equation (3-39*a*). As discussed above, if all the coefficients $P_r(x)$ are regular at $x = c$, that is, if they can be expanded in power series, c is an ordinary point of the differential equation.

If c is not an ordinary point but if $(x - c)^r P_r(x)$ is regular for all values of r, then $x = c$ is a regular singularity of the differential equation. Otherwise $x = c$ is an essential singularity. Sometimes these two kinds of singularity are called regular points and irregular points, respectively.

In accordance with the above criteria, the singularities in Legendre's equation (3-42) at $x = \pm 1$ are regular. The singularities in equation (3-39*a*) are also regular.

Problem 38. Show that Bessel's equation,

$$x^2\frac{d^2y}{dx^2} + x\frac{dy}{dx} + (x^2 - n^2)y = 0 \tag{3-48}$$

has a regular singularity at $x = 0$.

The formal procedure for finding a power-series solution in the neighborhood of a regular point is very similar to that used at an ordinary point. There are two principal differences.

1. It is not, in general, true that the various series solutions begin with the terms a_0, a_1x, a_2x^2, etc., as is the case at an ordinary point. The term with which each series begins must be determined from a study of the recursion formula, and its exponent may be negative and may be fractional.

2. It is not always possible to find enough linearly independent power-series solutions to make up a general solution. In this case other methods must be used, but they will not be discussed here.

The power-series method can be illustrated by applying it to find a solution of Bessel's equation in the neighborhood of the regular singularity at $x = 0$.

Let

$$y = x^\alpha \sum_{s=0}^{\infty} a_s x^s = \sum_{s=0}^{\infty} a_s x^{s+\alpha} \tag{3-48a}$$

This differs from the form used at an ordinary point by the presence of the factor x^α. Not only the coefficients a_s, but also the exponent α must be determined by substitution in the differential equation. To make the value of α unique, it is assumed that $a_0 \neq 0$ so that α is the exponent of the first term in the series.

Substitution of the series into the differential equation leads to

$$x^\alpha \sum_{s=0}^{\infty} a_s\{[(s + \alpha)^2 - n^2]x^s + x^{s+2}\} = 0 \tag{3-48b}$$

Equating to zero the coefficients of x^0, x^1, x^2, . . . , x^s in the summation leads to

$$\left.\begin{aligned} a_0(\alpha^2 - n^2) &= 0 \\ a_1[(\alpha + 1)^2 - n^2] &= 0 \\ a_2[(\alpha + 2)^2 - n^2] + a_0 &= 0 \\ a_s[(\alpha + s)^2 - n^2] + a_{s-2} &= 0 \end{aligned}\right\} \tag{3-48c}$$

The first of these equations is called the *indicial equation.* The last is the recursion formula. The indicial equation can be obtained from the recursion formula if the coefficients with negative subscripts are understood to be zero.

Since $a_0 \neq 0$, the indicial equation is a condition on α and shows that $\alpha = \pm n$. These values of α are called the exponents of the equation at the point $x = 0$. Except in the special case when $\alpha^2 = (\alpha + 1)^2$, the second equation requires that $a_1 = 0$. Even in this special case, a_1 may be taken to be zero. The third equation gives a_2 in terms of a_0. The recursion formula then permits the evaluation of all the coefficients in terms of a_0 and the value of α. Since there are two values of α, this procedure leads to two solutions, unless the recursion formula breaks down for one of the values.

The usual notation for the solutions of Bessel's equation labels them $J_{\pm n}(x)$, where the subscript corresponds to the lowest power of x in the expansion. When $n = \frac{1}{2}$, the two exponents are $\pm\frac{1}{2}$. If we take $a_0 = (2/\pi)^{\frac{1}{2}}$ to agree with the conventional definition,

$$J_{\frac{1}{2}}(x) = \left(\frac{2x}{\pi}\right)^{\frac{1}{2}} \left(1 - \frac{x^2}{6} + \frac{x^4}{120} - \cdots\right) \qquad (3\text{-}48d)$$

Similarly,

$$J_{-\frac{1}{2}}(x) = \left(\frac{2}{\pi x}\right)^{\frac{1}{2}} \left(1 - \frac{x^2}{2} + \frac{x^4}{24} - \cdots\right) \qquad (3\text{-}48e)$$

The general solution is then

$$y = AJ_{\frac{1}{2}}(x) + BJ_{-\frac{1}{2}}(x) \qquad (3\text{-}48f)$$

In the case of Bessel's equation with $n = \frac{1}{2}$, it is possible to form a general solution out of the two power series based on the two exponents of the equation. In general, this is not possible when the two exponents differ by an integer, and in such cases other methods must be used for determining a second solution.

Problem 39. Show that if $n = 2$ the power-series method will not give both solutions of Bessel's equation.

Problem 40. Show that

$$J_{-\frac{1}{2}}(x) = \left(\frac{2}{\pi x}\right)^{\frac{1}{2}} \cos x \qquad \text{and} \qquad J_{\frac{1}{2}}(x) = \left(\frac{2}{\pi x}\right)^{\frac{1}{2}} \sin x$$

and show that these forms satisfy Bessel's equation.

Problem 41. Show that $x = 0$ is a regular point of

$$\frac{d^2y}{dx^2} - \left(\frac{1}{4} - \frac{5}{36x^2}\right)y = 0$$

and find the power-series solutions about this point.

Problem 42. Show that, if the substitution of (3-48a) is made at an ordinary point of an equation such as (3-39), the method of procedure given above will produce both solutions.

CHAPTER IV

MECHANICS OF VIBRATING PARTICLES

1. Damped Vibrations.—There are many problems, in both mechanics and electromagnetism, that must be treated by means of the equations for vibrating particles. The mechanical problem is that of a particle attracted toward a fixed point with a force proportional to the distance from it and whose motion is at the same time opposed by a force proportional to the velocity. The mechanical problem is illustrated by the simple pendulum when the air resistance is considered; the electrical case is that of a circuit consisting of a capacitance, an inductance, and a resistance. The differential equation and the forms of the solution are the same for these two problems and for many other problems. For definiteness, consider the mechanical problem. The differential equation of motion is

$$m\frac{d^2x}{dt^2} + q\frac{dx}{dt} + kx = 0$$

where m is the mass of the particle, k is the force per unit displacement from the center, and q is the force per unit velocity. When k is positive, the force is in such a direction as to oppose the displacement and to restore the particle to an equilibrium position. When q is positive, the force associated with the velocity is in the direction to oppose it. Only positive values of these constants occur in simple cases. If k is negative, there is no position of equilibrium and no oscillation. If q is negative, the motion is not damped in the usual sense since the faster the particle is moving the more it is urged along. A negative q will not be expected in a simple mechanical oscillator, but it may occur in more complicated mechanical problems and in some electrical circuits.

The equation of motion is more conveniently written

$$\frac{d^2x}{dt^2} + p\frac{dx}{dt} + rx = 0 \tag{4-1}$$

where $p = q/m$ and $r = k/m$. By the methods of the previous chapter, the solution can be written at once. It is

$$x = A_0e^{\alpha_1 t} + B_0e^{\alpha_2 t} \tag{4-2}$$

where

$$\alpha_1 = -\frac{p}{2} + \frac{1}{2}(p^2 - 4r)^{\frac{1}{2}} \qquad \alpha_2 = -\frac{p}{2} - \frac{1}{2}(p^2 - 4r)^{\frac{1}{2}}$$

The nature of the motion represented by this solution depends upon whether α_1 and α_2 are real, imaginary, or complex numbers. This in turn depends upon whether $p^2 - 4r$ is zero, negative, or positive and whether p is zero or positive. Only cases in which p is not negative will be considered here.

a. Underdamped Motion.—When $p^2 - 4r < 0$, the motion is said to be *underdamped.* In this case the restoring force, which tends to maintain the vibration, is dominant; and the particle executes oscillations about its position of equilibrium in spite of the damping force, which tends to stop it. Let $\omega_0 = \frac{1}{2}(4r - p^2)^{\frac{1}{2}}$. Then equation (4-2) can be written in three forms.

$$x = e^{-pt/2}(A_1e^{i\omega_0 t} + B_1e^{-i\omega_0 t}) \tag{4-2a}$$

$$x = e^{-pt/2}(A \cos \omega_0 t + B \sin \omega_0 t) \tag{4-2b}$$

$$x = Ce^{-pt/2} \cos (\omega_0 t - \varphi) \tag{4-2c}$$

Each of these represents an oscillation with an exponentially diminishing amplitude as shown in Fig. 4-1.

The arbitrary constants A_1 and B_1 in (4-2*a*), A and B in (4-2*b*), or C and φ in (4-2*c*) are so chosen as to fit the desired conditions of any particular motion. When the complex form of equation (4-2*a*) is used, the constants will turn out such that, even though the expression is written in complex notation, its value is real. This frequently means that the constants themselves are complex. The specification of the position and velocity, x and $\dot{x}$, at any given time, for example, will suffice

to fix these constants. The more characteristic properties of the motion, however, are fixed by the constants p and r that occur in the differential equation. Various combinations of these constants give useful characteristics.

Even though the motion of a damped oscillator is not strictly periodic, since the amplitude is decreasing, it is convenient to

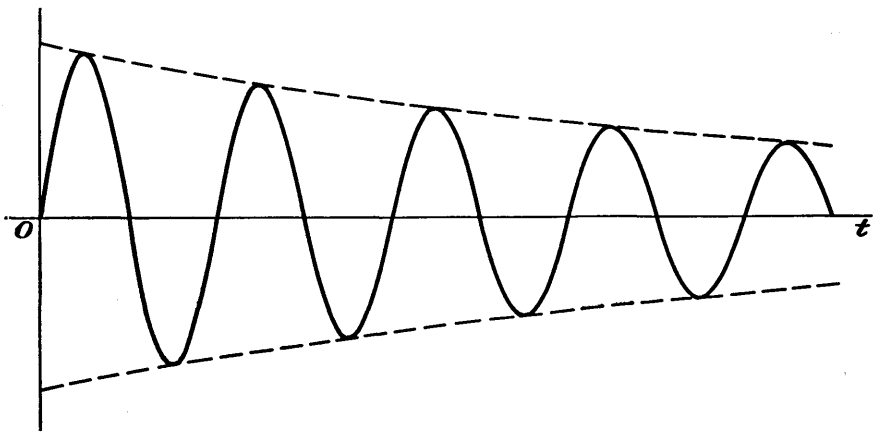


FIG. 4-1.—The displacement as a function of the time for an underdamped oscillator. The broken line represents the exponential decrease.

define a *period*, which is the time between two successive passages through zero in the same sense. This turns out to be

$$T = \frac{2\pi}{\omega_0} = \frac{4\pi}{(4r - p^2)^{\frac{1}{2}}} \tag{4-3}$$

The period characterizes the oscillatory nature of the motion.

The damping of the oscillations is described in several ways. The *logarithmic decrement* is the logarithm of the ratio of two successive maxima of the same sign. It is equal to $p/2$ times the period so that

$$\delta = \frac{pT}{2} = \frac{\pi p}{\omega_0} = \frac{2\pi p}{(4r - p^2)^{\frac{1}{2}}} \tag{4-4}$$

It is also possible to define a *modulus of decay* as the time during which the amplitude decreases to $1/e$ of its initial value. The modulus of decay is obviously equal to $2/p$. A third constant is

used frequently in electrical circuits and acoustic devices and is known merely as Q. It is given by

$$Q = \frac{\omega_0}{p} \tag{4-5}$$

A circuit with high Q is only slightly damped. Its energy dissipation is low, and its resonance is sharp, as will be shown later.

A damped oscillator is not a conservative system. Initially it has a certain potential and kinetic energy, but this is eventually used up in overcoming the damping resistance. If one wished to include the sources of the damping resistance in the system under consideration, there would be conservation of energy in the whole system; but this is not usually convenient. It is usually most convenient to regard the work done against the damping force as simply a dissipation of energy.

The energy of the oscillator is the sum of the potential and the kinetic energies,

$$W = \frac{mr}{2} x^2 + \frac{m}{2} \dot{x}^2 \tag{4-6}$$

When the damping is neglected, $p = 0$, this sum is constant, and a decrease in x^2 is accompanied by an increase in $\dot{x}^2$. When $p > 0$, equation (4-2c) can be used to express the energy as a function of the time. This leads to

$$W = \frac{m}{2} C^2 e^{-pt} \left[\omega_0^2 + \frac{p^2}{2} \cos^2 (\omega_0 t - \varphi) + p\omega_0 \sin (\omega_0 t - \varphi) \cos (\omega_0 t - \varphi) \right] \tag{4-7}$$

and shows that the energy does not decrease steadily, as would be suggested by the exponential alone, but in a manner related to the oscillations. Energy is dissipated only when the particle is moving. The rate of dissipation falls to zero at the extreme excursions where the particle stops and starts back.

Problem 1. Show that equation (4-2) is a solution of the differential equation (4-1). Show also that (4-2a), (4-2b), and (4-2c) are

different forms of (4-2) for the underdamped case, and express the constants C and φ in terms of A and B.

Problem 2. Evaluate the arbitrary constants of (4-2*b*) in terms of the initial position and velocity of the particle. Sketch the motion as a function of the time when the initial velocity is zero.

Problem 3. Show that the time between successive maxima of displacement is T.

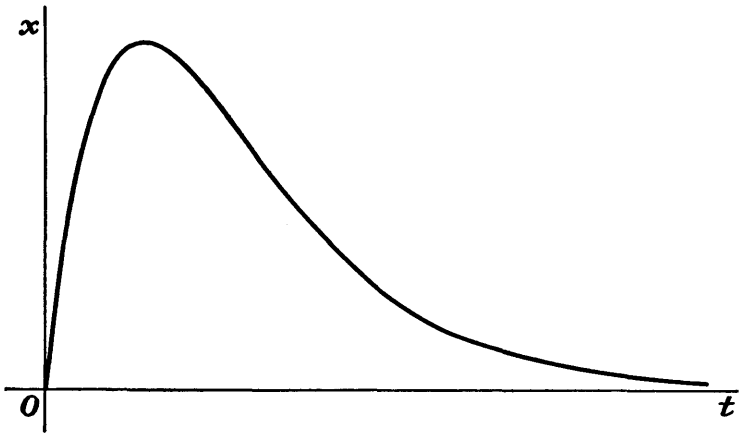


FIG. 4-2.—The displacement as a function of the time for a critically damped oscillator.

Problem 4. Work out the rate of dissipation of energy against the damping force of an underdamped oscillator, and show that it is equal to the rate of decrease of total energy of the oscillator.

b. Critically Damped Motion.—When $p^2 - 4r = 0$, the motion is said to be *critically damped.* In the critically damped case the auxiliary equation has equal roots, and the solution can be written

$$x = (A + Bt)e^{-pt/2} \tag{4-8}$$

Figure 4-2 illustrates this motion for the case in which the particle starts from the origin with an initial velocity. It attains a maximum displacement and then approaches the equilibrium position asymptotically. Under these conditions the curve never crosses the axis. If the particle starts with an initial displacement and a sufficiently great velocity toward the origin, it will cross the axis once and then return to it asymptotically from the other side.

Problem 5. Evaluate the constants in equation (4-8) in terms of the initial conditions. Find the time and magnitude of the maximum displacement when the particle starts from the origin with an initial velocity.

c. Overdamped Motion.—When $p^2 - 4r > 0$, the motion is said to be *overdamped.* This is the case in which the damping force is dominant, and in extreme cases the restoring force can be neglected altogether. The two roots of the auxiliary equation, α_1 and α_2, are real and negative, and the solution can be used directly in the form of equation (4-2).

Problem 6. Consider an overdamped oscillator that starts from the origin with an initial velocity. Find the time at which the maximum displacement occurs, and find its value.

Problem 7. An overdamped oscillator has an initial displacement x_0 and an initial velocity $\dot{x}_0$. For what values of x_0 and $\dot{x}_0$ will the particle pass the origin before coming to rest?

Problem 8. Write and solve the differential equation for a damped motion in which the elastic restoring force is neglected entirely. Show how this solution is related to equation (4-2).

2. Forced Vibrations.—In most cases of practical interest, the vibrating system is acted on by an outside force. This is the case of vibrations in the mounting of a machine containing an unbalanced rotating part. It is also the case of excitation of an electrical resonant circuit by electromagnetic waves. In these two examples, the external force is periodic in time, and this, in fact, constitutes the most important case because of the phenomenon of resonance. When the period of the oscillating system and the period of the external periodic force are nearly the same, the force is most effective in stimulating oscillations. The amplitude of the motion then builds up until the work done by the force is dissipated against the damping resistance. When the periods of the force and the oscillation differ widely, the force does little work, for part of the time the motion is in the direction of the force and part of the time against it.

If the external force acting on the oscillator is $mF(t)$, the equation of motion is

$$\frac{d^2x}{dt^2} + p\frac{dx}{dt} + rx = F(t) \tag{4-9}$$

$F(t)$ is represented as a function of the time only, for this is the only case that is properly called a forced oscillation. If the force is a function of position or of velocity, the motion is of a completely different character.

Equation (4-9) is a nonhomogeneous linear differential equation. The general solutions of equations of this type were discussed in the previous chapter. The interest here lies in investigating the properties of the solution under various circumstances.

Problem 9. Show that a constant force acting on an oscillator is equivalent to a displacement of the origin.

Problem 10. Consider an oscillator that starts from the origin with an initial velocity $\dot{x}_0$ and is subject to an external force $mF(t)$. $F(t) = 0$ for $0 < t < t_1$, and $F(t) = F_0$ for $t > t_1$. Find the motion.

Problem 11. Find the motion of an oscillator acted on by a force proportional to the time. Evaluate the arbitrary constants if the velocity and the displacement are zero at $t = 0$.

a. Harmonic Forces.—Most of the important properties of forced oscillators can be derived by a consideration of sinusoidal forces. This is because the forces in many practical cases can be closely approximated by a sine or cosine function of the time. In addition, as will be shown later in a discussion of Fourier series, any periodic function of the time can be expressed as a sum of sine and cosine terms.

If $F(t)$ in equation (4-9) is given by $F_0 \sin \omega t$, a particular integral of the equation is given by

$$x = \frac{F_0 \sin (\omega t - \epsilon)}{[(r - \omega^2)^2 + p^2\omega^2]^{\frac{1}{2}}} \tag{4-10}$$

with

$$\tan \epsilon = \frac{p\omega}{r - \omega^2} \tag{4-10a}$$

where ϵ is between 0 and π. This shows that the motion of the oscillator under the influence of the harmonic force is a sinusoidal motion with the frequency of the force. The motion is out of phase with the force by the amount of the *phase lag* ϵ, and the amplitude depends upon the relationship between the frequency of the force and the constants of the oscillator.

The amplitude for very small and very large values of ω is easily understood from physical considerations or from a consideration of the differential equation (4-9). When $\omega \to 0$, both the second derivative and the first derivative become negligible compared with the term rx. Hence the equation is effectively

$$rx = F_0 \sin \omega t \tag{4-11}$$

and the solution is clearly

$$x = \frac{F_0}{r} \sin \omega t \tag{4-11a}$$

This same result is obtained when ω is set equal to zero in equations (4-10) and (4-10a). Physically it represents a displacement such that the restoring force just balances the slowly varying external force.

For a very rapidly varying force, the second derivative is dominant, so that the differential equation approaches

$$\frac{d^2x}{dt^2} = F_0 \sin \omega t \tag{4-12}$$

of which a particular integral is

$$x = -\frac{F_0}{\omega^2} \sin \omega t = \frac{F_0}{\omega^2} \sin (\omega t - \pi) \tag{4-12a}$$

Here again this agrees with equations (4-10) and (4-10a) when $\omega \to \infty$. Physically it means the displacement is so small that the restoring force is unimportant; the velocity stays so small that the frictional force is negligible; and the rapidly varying external force is just equal to the mass multiplied by the acceleration.

It is also possible to give a physical interpretation to the

case in which $\omega = \sqrt{r}$. In this case the acceleration is just equal to the restoring force divided by the mass, so that the particle moves as though there were no damping whatever. The frictional resistance is then balanced by the external force, so that the amplitude can be obtained from the differential equation

$$p\frac{dx}{dt} = F_0 \sin \omega t \tag{4-13}$$

A particular integral of this equation is

$$x = -\left(\frac{F_0}{\omega p}\right)\cos \omega t = \left(\frac{F_0}{\omega p}\right)\sin\left(\omega t - \frac{\pi}{2}\right) \tag{4-13a}$$

This is the form taken by equation (4-10) when $\omega^2 = r$.

Figure 4-3 shows the amplitude as a function of ω for several values of the damping coefficient, and Fig. 4-4 shows the phase lag ϵ. The phenomenon of resonance is shown by the sharp maximum in the amplitude when the damping is small and by the corresponding rapid change in ϵ as ω passes through $\sqrt{r}$. This maximum occurs in the neighborhood of $\sqrt{r}$, which is the angular frequency of the undamped oscillator, but not exactly at this frequency. Neither does it occur at the frequency of the free damped oscillator ω_0. However, these differences are small when the damping is small.

Problem 12. Find the frequency ω_m at which the amplitude of a forced oscillator is a maximum.

Problem 13. Find the frequency at which the maximum velocity of a forced oscillator has its greatest value.

Problem 14. Show that the frequency ω_c, at which the amplitude of a forced oscillator has $1/\sqrt{2}$ its maximum value, satisfies the relationship

$$\omega_c^2 = \omega_m^2 \pm p\omega_0 \tag{4-14}$$

Although equation (4-10) is a particular integral of equation (4-9), it is not the general solution. It cannot be fitted to arbitrary initial conditions. The general solution is obtained by adding the general solution of the homogeneous equation. Hence for an oscillator subject to a sinusoidal force

$$x = A_0 e^{\alpha_1 t} + B_0 e^{\alpha_2 t} + \frac{F_0 \sin(\omega t - \epsilon)}{[(r - \omega^2)^2 + p^2\omega^2]^{\frac{1}{2}}} \tag{4-15}$$

Since the real parts of α_1 and α_2 are negative, the contribution of the general solution eventually disappears and the steady motion is that of the particular integral. The contribution of

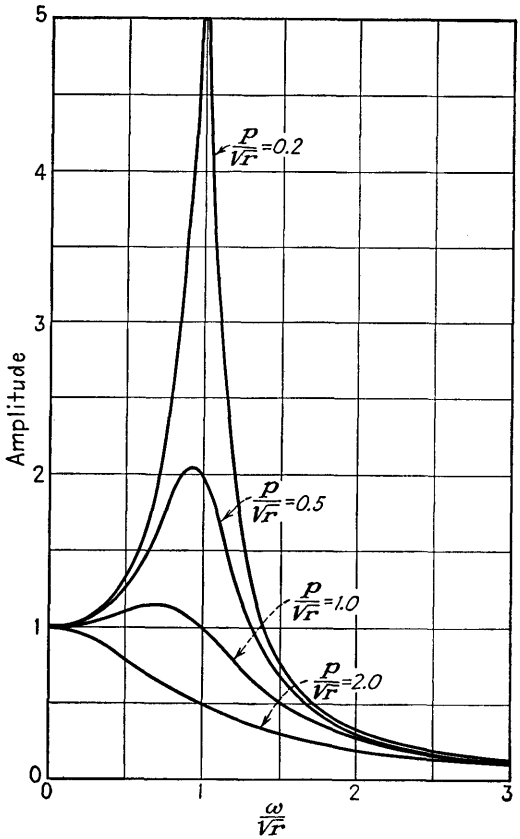


FIG. 4-3.—The amplitude of a damped oscillator, driven by a sinusoidal driving force, as a function of the frequency of the driving force.

the complementary function is hence called a *transient.* It is often important to know the transient behavior of an oscillator, since the time necessary to build up the oscillation to its steady state may be of importance.

To get the maximum response at resonance it is desirable to make p as small as possible, but with a small damping the transient dies out slowly and some time is required for the

steady state to build up. At the frequency of maximum amplitude

$$\frac{x_{\max}}{F_0} = \frac{1}{p\omega_0} \tag{4-16}$$

At the start of the motion the transient term is just equal and opposite to the term describing the steady state, so that the resultant is zero. As the transient dies out, the resultant builds

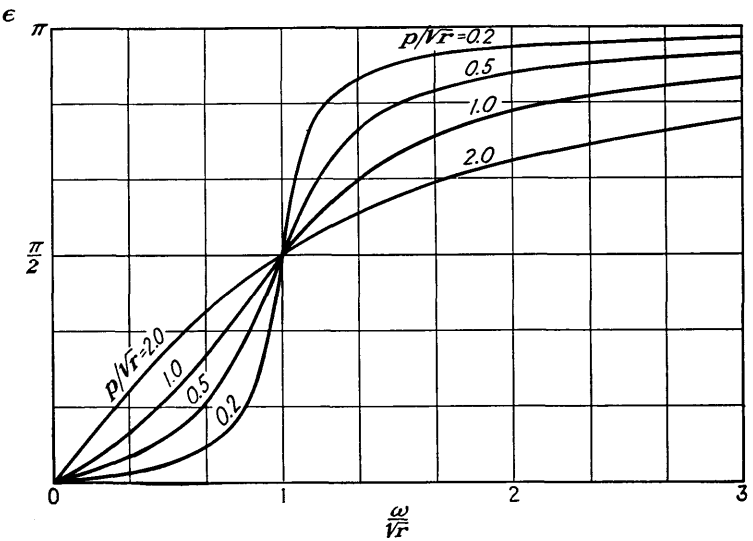


FIG. 4-4.—The phase lag of a forced oscillator.

up, so that the rate of growth of the oscillation is just given by the rate of decay of the transient term of the general solution. The time necessary for the transient to decrease to e^{-K} of its original magnitude is

$$\tau = \frac{2K}{p} \tag{4-16a}$$

Hence the average rate of growth of the amplitude, is

$$\frac{x_{\max}}{F_0\tau} = \frac{1}{2K\omega_0} \tag{4-16b}$$

and is independent of the damping, except as p is contained in ω_0.

Problem 15. Evaluate the arbitrary constants of equation (4-15) when the particle starts from rest at the origin.

b. Energy Absorption.—In the steady state, the external force is doing work on the oscillator at the same average rate as that at which the oscillator is working against the damping force. During each cycle the work done by the external force is just equal to the work done against the resistance. The work is not always done directly against the damping resistance, however. It is stored as potential and kinetic energy of the oscillator during part of the cycle and dissipated during another part. The energy of the oscillator is not constant during the cycle, except when the force has the frequency of the undamped oscillator. Usually the oscillator, even though it is moving with simple-harmonic motion, does not have a constant energy because its frequency does not correspond to its restoring force and mass.

Problem 16. Show that the energy of a forced oscillator is constant if the force has a frequency corresponding to the free oscillator without damping.

Problem 17. Compute the energy of a forced oscillator as a function of the time.

Problem 18. Compute the rate at which work is done on a forced oscillator by the damping force.

c. Vibration Insulation.—It is often necessary to protect a sensitive instrument, such as a galvanometer or an interferometer, against vibrations from external sources. This can be done by suspending a heavy base on springs and possibly inserting some damping between the base and the support. Figure 4-5 indicates schematically the arrangement. In this figure only vertical motion is considered, and the instrument base is supported from the floor. The floor is then supposed to move vertically with a sinusoidal motion. The springs and the damping mechanism both operate by virtue of a difference between the displacements of the floor and the base. The object of the mounting, of course, is to reduce as far as possible the motion of the base.

Let x_0 be the vertical displacement of the floor from its equilibrium position, and let x be the corresponding displacement of the instrument base. Then the equation of motion is

$$m\frac{d^2x}{dt^2} = -q\left(\frac{dx}{dt} - \frac{dx_0}{dt}\right) - k(x - x_0) \tag{4-17}$$

For convenience let $x - x_0 = X$. Then

$$m\frac{d^2X}{dt^2} + q\frac{dX}{dt} + kX = -m\frac{d^2x_0}{dt^2} \tag{4-17a}$$

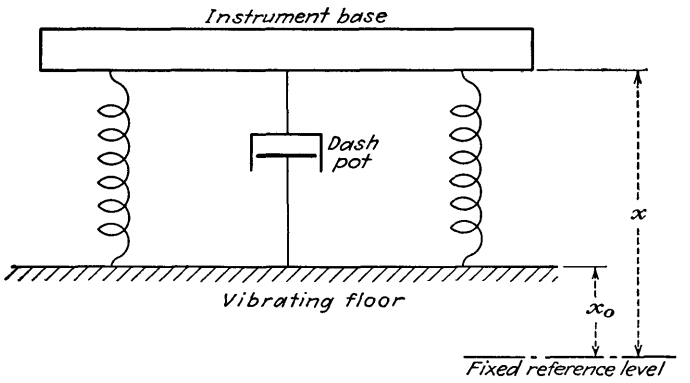


FIG. 4-5.—Schematic illustration of an instrument base mounting designed to isolate the mechanism from vibrations of the floor.

If $x_0 = A \sin \omega t$

$$\frac{d^2X}{dt^2} + p\frac{dX}{dt} + rX = A\omega^2 \sin \omega t \tag{4-18}$$

By the methods already discussed it can be shown that the steady-state solution of equation (4-18) leads to

$$\frac{x}{A} = \frac{[r^2 + p^2\omega^2]^{\frac{1}{2}}}{[(r - \omega^2)^2 + p^2\omega^2]^{\frac{1}{2}}} \sin (\omega t - \epsilon') \tag{4-19}$$

The object of the design is to make x/A as small as possible. A detailed study of equation (4-19) shows that r should be made as small as possible to protect against a given frequency ω. In fact $x/A > 1$ if $\omega^2 < 2r$. For large ω the damping is of little use and tends to increase the transmitted motion. For $\omega^2 < 2r$ the damping is helpful.

Problem 19. Show that equation (4-19) follows from equation (4-18).

Problem 20. Show that, for $\omega^2 = 2r$ in equation (4-19), $x/A = 1$ and the vibration mounting is useless.

Problem 21. Show that for $\omega^2 = r$ the motion of the base is greater than that of the floor but that the damping is helpful in keeping it from attaining very large values.

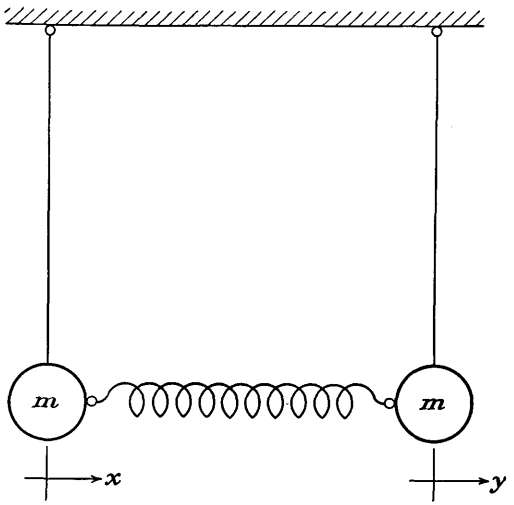


FIG. 4-6.—Schematic drawing of two coupled pendulums.

Problem 22. Show that, for $\omega^2 > 2r$, $x/A < 1$ and that damping is detrimental.

3. Coupled Oscillators.—If two or more vibrating systems are connected in such a way that they influence each other, the differential equations of motion are simultaneous linear equations with constant coefficients. The method of treatment in this case can be illustrated by the following problem:

Consider two simple pendulums of length l and mass m, Fig. 4-6. Let the bobs be connected by a massless spiral spring whose force constant is a. This fiction of massless springs is customary in mechanical problems and simply means that the inertia of the spring is not to be taken into account. The spring serves merely to represent a force between the two masses, which is proportional to the difference between their

actual separation and their separation in the normal rest position. Let x be the displacement of the left-hand pendulum from its position of rest, and let y be the displacement of the other from its position of rest. Both displacements are measured to the right and are assumed to be so small that the motion of a pendulum alone can be treated as simple harmonic. The natural length of the spring is equal to the distance between the vertical positions of rest. Then the equations of motion are

$$\left.\begin{aligned} m\frac{d^2x}{dt^2} &= -\frac{mgx}{l} - a(x-y) \\ m\frac{d^2y}{dt^2} &= -\frac{mgy}{l} + a(x-y) \end{aligned}\right\} \tag{4-20}$$

To find a solution of these equations, let $x = Ae^{i\omega t}$, and let $y = Be^{i\omega t}$. This substitution represents a motion in which both masses move with simple-harmonic motion and with the same frequency. Advantage has been taken of the fact that the motion is undamped, and the exponents have been written so as to be pure imaginaries when ω is real. The substitution of these forms into the differential equations gives a pair of algebraic equations for the determination of A and B.

$$\left.\begin{aligned} -\left(m\omega^2 - \frac{mg}{l} - a\right)A - aB &= 0 \\ -aA - \left(m\omega^2 - \frac{mg}{l} - a\right)B &= 0 \end{aligned}\right\} \tag{4-21}$$

Both these equations must be satisfied if it is to be possible to find a solution of the form assumed, and they can be satisfied by values different from zero only if they are not contradictory. The two equations are compatible only if the determinant of the coefficients is equal to zero. This determinant contains the undetermined quantity ω; and, by a proper selection of the values of ω, it is possible to ensure the compatability of the equations. This condition then fixes the frequency of the vibration. The expansion of the determinant gives

$$\left(m\omega^2 - \frac{mg}{l}\right)^2 - 2a\left(m\omega^2 - \frac{mg}{l}\right) + a^2 - a^2 = 0 \tag{4-22}$$

The values of ω that satisfy this equation are

$$\omega_1^2 = \frac{g}{l} + \frac{2a}{m} \qquad \text{and} \qquad \omega_2^2 = \frac{g}{l} \tag{4-22a}$$

Including both the positive and the negative roots of ω^2, there are really four values of ω for which it is possible to write a solution in the form assumed. For each value of ω, the corresponding ratio between the coefficients A and B is given by equations (4-21). For $\omega = \pm\omega_1$, $A = -B$; while for $\omega = \pm\omega_2$, $A = B$. Hence the general solution of equations (4-20) is

$$\left.\begin{aligned} x &= A_1 e^{i\omega_1 t} + A_2 e^{-i\omega_1 t} + A_3 e^{i\omega_2 t} + A_4 e^{-i\omega_2 t} \\ y &= -A_1 e^{i\omega_1 t} - A_2 e^{-i\omega_1 t} + A_3 e^{i\omega_2 t} + A_4 e^{-i\omega_2 t} \end{aligned}\right\} \tag{4-23}$$

In this solution there are the four arbitrary constants that are to be expected in the solution of two second-order differential equations so that these solutions can be adapted to an arbitrary set of initial conditions.

It is important to understand the significance of equation (4-22). Only when this equation is satisfied is it possible to satisfy simultaneously the two equations (4-21), and only when these two are satisfied is it possible for the assumed exponential forms to satisfy the differential equations. Equation (4-22) gives the values of ω for which a solution can be found in the form assumed. In the example treated, there are four usable values and consequently four different solutions. Each of these is a particular solution of the problem, while the sum of them, each multiplied by an arbitrary constant, is the genera solution.

When the energy of this system of coupled pendulums is written down, care must be taken to include the potential energy of the coupling spring in the proper way. The kinetic energy is simply the sum of the energies of the two masses.

$$T = \frac{m}{2}\dot{x}^2 + \frac{m}{2}\dot{y}^2 \tag{4-24}$$

The potential energy can be visualized most simply as consisting of two terms representing the potential energy of each

particle in the gravitational field and a term representing the energy stored in the spring. Hence

$$V = \frac{mg}{2l} x^2 + \frac{mg}{2l} y^2 + \frac{a}{2} (x - y)^2 \qquad (4\text{-}24a)$$

A more formal but still instructive way to arrive at the potential energy is to consider the work necessary to move the system from its position of equilibrium to its final position. First hold the right-hand pendulum fixed and displace the left hand one by the amount x. The work done is

$$W_1 = \int_0^x \left(\frac{mg}{l} x + ax\right) dx = \frac{mg}{2l} x^2 + \frac{a}{2} x^2$$

Then, holding the left-hand mass at x, move the right-hand one to y. The work is

$$W_2 = \int_0^y \left[\frac{mg}{l} y + a(y - x)\right] dy = \frac{mg}{2l} y^2 + \frac{a}{2} y^2 - axy$$

The sum of these two terms is equal to the potential energy V of equation (4-24a).

Problem 23. Evaluate the constants in equations (4-23) in terms of the initial positions and velocities.

Problem 24. Express the solution (4-23) in trigonometric form.

Problem 25. Find the initial conditions for which A_1 and A_2 are zero and also those for which A_3 and A_4 are zero. A vibration under either of these conditions is a "normal vibration."

Problem 26. In the case where a/m is small compared with g/l, express the motion of each pendulum as a simple vibration with a variable amplitude.

Problem 27. Treat the case in which the masses of the two pendulums are different.

4. Normal Coordinates.—As was shown in Prob. 25, it is possible to set this system of coupled pendulums vibrating in such a way that the vibration takes place with one frequency only. From equations (4-23) it is possible to show that

$$\left.\begin{aligned} (x + y) &= X = 2A_3 e^{i\omega_2 t} + 2A_4 e^{-i\omega_2 t} \\ (x - y) &= Y = 2A_1 e^{i\omega_1 t} + 2A_2 e^{-i\omega_1 t} \end{aligned}\right\} \qquad (4\text{-}25)$$

Each of the quantities X and Y varies harmonically, but with its individual frequency. When $X = 0$ and Y is vibrating, there is one normal vibration excited; when $Y = 0$ and X is vibrating, the other normal vibration is excited. In general, both are excited simultaneously. The quantities X and Y are called the *normal coordinates* of the system, and they are just as satisfactory as the original x and y for specifying the configuration. These normal coordinates have in addition, however, the advantage that they are entirely independent of one another in their vibration.

In general, normal coordinates are linear combinations of the ordinary coordinates. In the simple and symmetrical case of the two identical pendulums the normal coordinates are simply the sum and the difference as shown in equations (4-25). In other cases such as that of Prob. 27 the linear combinations are less simple. In that problem

$$\left.\begin{aligned} X_1 &= x_1 + \frac{m_2}{m_1} y_1 \\ Y_1 &= x_1 - y_1 \end{aligned}\right\} \tag{4-25a}$$

constitute the normal coordinates. In these simple cases such coordinates can be obtained from the general solutions by inspection.

Problem 28. Make the transformation to the coordinates X and Y in the differential equations (4-20). Write the solutions of the transformed equations.

Problem 29. Express in terms of the normal coordinates the energy of the system of two coupled pendulums shown in Fig. 4-6.

Problem 30. Treat the problem of two coupled pendulums when each mass moves against a frictional force that is proportional to the velocity.

Problem 31. Three equal masses are confined in a frictionless tube, which is in a horizontal position. The masses are separated from each other and from the ends of the tube by four springs of the same length and the same force constants. Find the motion of the masses.

Problem 32. Find the normal coordinates for the above problem.

From the above treatment, the following properties of normal coordinates can be tabulated:

1. Normal coordinates are coordinates in which the equations of motion take the form of a set of linear differential equations with constant coefficients and in which each equation contains one dependent variable only.

2. Normal coordinates are coordinates that are independent of each other in the sense that one can be set into vibration while the others remain at rest. This fact can, of course, be derived from the general solution of the equations mentioned in item 1. A vibration in which only one normal coordinate is vibrating is called a *normal mode of vibration.*

3. Normal coordinates are coordinates in which the total energy of the (undamped) system can be expressed as a sum of the squares of the coordinates multiplied by constant coefficients and a sum of the squares of the first derivatives of the coordinates multiplied by constant coefficients.

Normal coordinates do not otherwise differ from arbitrarily chosen coordinates and can be used for the description of the system in the same manner as the original ones. When there are only two coordinates, it is sometimes convenient to plot the configuration of the system as a point in an x-y plane. This point then gives the value of the two coordinates and so completely defines the system. The motion of the system can then be described by an orbit in this plane.

5. Oscillations under an External Force.—In systems of two and more degrees of freedom, an external force can be applied in a variety of ways. In the problem of the two pendulums if the same periodic force is applied to both of them the normal coordinate X will be set into vibration. If the force is applied to the two pendulums in opposite directions, the coordinate Y will be excited. If the force is applied to one pendulum only, both normal modes of vibration will be stimulated.

As an illustration, consider the case in which both pendulums are subject to a damping resistance equal to mp times the velocity, and let a force $mF_0 \sin \omega t$ act on the left-hand pendulum. The problem can be treated in several ways. One way is to

write down the differential equations and substitute an exponential. Perhaps a simpler way is to transform the differential equations directly to normal coordinates. This leads to

$$\left.\begin{aligned} \frac{d^2X}{dt^2} + p\frac{dX}{dt} + \frac{g}{l}X &= \frac{F_0}{2}\sin \omega t \\ \frac{d^2Y}{dt^2} + p\frac{dY}{dt} + \left(\frac{g}{l} + \frac{2a}{m}\right)Y &= \frac{F_0}{2}\sin \omega t \end{aligned}\right\} \tag{4-26}$$

These equations are of the form already treated, and the solutions can be written down at once.

$$\left.\begin{aligned} X &= \frac{\frac{1}{2}F_0 \sin(\omega t - \epsilon)}{\left[\left(\frac{g}{l} - \omega^2\right)^2 + p^2\omega^2\right]^{\frac{1}{2}}} \\ Y &= \frac{\frac{1}{2}F_0 \sin(\omega t - \epsilon')}{\left[\left(\frac{g}{l} + \frac{2a}{m} - \omega^2\right)^2 + p^2\omega^2\right]^{\frac{1}{2}}} \end{aligned}\right\} \tag{4-27}$$

with

$$\tan \epsilon = \frac{p\omega}{(g/l) - \omega^2} \qquad \text{and} \qquad \tan \epsilon' = \frac{p\omega}{(g/l) + (2a/m) - \omega^2}$$

One normal coordinate has its resonance at one frequency and the other at a different frequency.

The motion of either pendulum can be obtained by taking the sum or difference of the normal coordinates. Each individual mass shows resonance at two frequencies. The magnitude of the oscillation as well as the phase can be worked out from equations (4-27).

Problem 33. Show that equations (4-26) follow from the statement of the problem.

CHAPTER V

CALCULUS OF VARIATIONS

The Newtonian equations of motion, as postulated in Chap. II, have a very simple form; the force along each coordinate axis is represented by a single letter, and its dependence on the coordinates is not specified. When the functional form of the forces is specified in order to give the equations of motion for a specific problem, a good deal of complication may result, and most problems can best be handled by introducing a coordinate system particularly adapted to the situation. Frequently the direct transformation of the equations of motion to the new coordinate system is a matter of some difficulty. Even the transformation to plane polar coordinates, carried out in Chap. II, is laborious. If, however, the equations of motion are written in the form of a *variation principle*, which is independent of the particular coordinate system used, the transformation to various systems can be much simplified. This method of writing the equations of motion will be considered in the next chapter. The present chapter will deal with some of the more elementary methods of the calculus of variations that must be used in dealing with variation principles.

1. The Variation Problem.—The standard problem of the calculus of variations is that of finding the form of the function $y = y(x)$ such that a given integral

$$I = \int_{x_1}^{x_2} \Phi\left(y,x,\frac{dy}{dx}\right) dx \tag{5-1}$$

shall have a maximum or a minimum value. The function Φ is given, and in many cases the limits x_1 and x_2 are prescribed. The only thing that can be changed in the attempt to make I larger or smaller is the form of the function y. In some cases the limits of integration may not be fixed, and then they, too, can be varied.

The following three examples of variation problems will serve to illustrate the methods of the calculus of variations and the types of problems that can be treated.

a. What is the shortest line between two points in a plane? To put this problem in the standard form indicated above, take x and y axes in the plane, and let (x_1,y_1) and (x_2,y_2) be the two points between which the line is to be drawn. The line connecting the points will be represented by the function $y = y(x)$. The length of the line between the two points is then given by the integral

$$L = \int_{(x_1,y_1)}^{(x_2,y_2)} (dx^2 + dy^2)^{\frac{1}{2}} = \int_{x_1}^{x_2} \left[1 + \left(\frac{dy}{dx}\right)^2\right]^{\frac{1}{2}} dx \qquad (5\text{-}2)$$

The problem is to find the form of y as a function of x that will make the integral L a minimum.

b. Another classical problem is that of the brachistochrone. Given two points in a vertical plane, the coordinates of which are (x_1,y_1) and (x_2,y_2), what is the curve along which a particle will slide from one to the other in the shortest time? For simplification let the upper point be at the origin of coordinates so that $x_1 = y_1 = 0$, and let the x axis be horizontal. Then the speed of the particle is a function of y only and is given by

$$v = (-2gy)^{\frac{1}{2}} \qquad (5\text{-}3)$$

when the initial velocity is assumed to be zero. The time occupied in passing over an element of the curve is

$$dt = \left(\frac{dx^2 + dy^2}{-2gy}\right)^{\frac{1}{2}}$$

and the whole time is

$$t = \int_0^{x_2} \left[\frac{1 + (dy/dx)^2}{-2gy}\right]^{\frac{1}{2}} dx \qquad (5\text{-}4)$$

The problem is to make this integral a minimum by finding the proper form for y as a function of x.

c. A modification of the brachistochrone problem involves a case in which the ordinate of the curve is not fixed at both limits of integration. A special case is the problem of finding

the curve down which a particle will slide in the shortest time from a given point to a given vertical line. Since the point at which the curve meets the vertical line is not specified, the end point of the curve is not fixed and may be changed in the attempt to find a curve for which the time is a minimum. The integral for this problem is the same as for the previous one. The limit of integration x_2 gives the position of the vertical line, but the value of y_2 must be determined as part of the solution.

The mathematical form of equations (5-2) and (5-4) in which these problems are expressed is valid only when a number of conditions are satisfied. In particular, the two limits x_1 and x_2 must not coincide. If they do, the axes should be selected in some other way. For a detailed discussion of these restrictions reference should be made to a standard treatise on the calculus of variations.

2. Extreme Values of Ordinary Functions.—The problem of finding the maxima and minima of functions of one or more independent variables is treated in the ordinary differential calculus. To determine whether a certain value of a single independent variable x_0 gives a maximum or minimum value to the function $y = y(x)$, the value of the function for $x = x_0$ is compared with the values of the function for neighboring values of the independent variable. To find the values of the function for neighboring values of x, it is convenient to use a Taylor's expansion. This gives

$$y(x) = y(x_0) + (x - x_0)y'(x_0) + \tfrac{1}{2}(x - x_0)^2 y''(x_0) + \cdots \quad (5\text{-}5)$$

If $y(x_0)$ is a maximum or a minimum value, the second term in this expansion will vanish; and this requires that the first derivative of y at the point x_0 shall vanish. The vanishing of the first derivative does not distinguish between a maximum and a minimum, nor is it a sufficient condition for either. It is, however, a necessary condition.

In dealing with a function of one real variable, there is no uncertainty as to what values of the independent variable must be considered. The independent variable can be only increased

or decreased, and the first derivative is uniquely defined. In dealing with a function of two or more independent variables, however, the situation is already a little more complicated. Such a function may have a maximum with reference to some values of the independent variables and at the same time have a minimum with respect to others. In such a case a function has a true, or absolute, maximum only when it is greater at a given point than at *any* other point in its neighborhood.

An example of this situation is the function

$$y(x,z) = (x - x_0)^2 - (z - z_0)^2 \tag{5-6}$$

The point $(x = x_0, z = z_0)$ is one at which the function has a stationary value; the first derivatives with respect to x and with respect to z are zero at this point and are zero no matter how the x and z axes are rotated. However, this is clearly neither a true minimum nor a true maximum. It is a minimum with respect to points on the line parallel to the x axis through z_0 and is a maximum with respect to points on the line parallel to the z axis through x_0. At all points along the lines

$$x - x_0 = \pm(z - z_0)$$

the function has the constant value zero. Furthermore, any point whose z coordinate is z_0 gives to the function a maximum value compared with those points having the same value of x but different values of z, but at no point does the function have a true maximum. This function serves to illustrate two points.

1. The vanishing of the first derivative in all directions is only a necessary and not a sufficient condition for the existence of a maximum or a minimum. At the point (x_0,y_0) the first derivative vanishes, but this point is neither a maximum nor a minimum. However, at this point the function may be said to have a stationary value, and the vanishing of the first derivative may be taken as both necessary and sufficient for the existence of such a value. It may be taken as the definition of a stationary value.

2. The vanishing of the first derivative with respect to only *one* of the independent variables is not sufficient even for the

existence of a stationary value. At all points along the line $x = x_0$, $\partial y/\partial x = 0$, but the function does not have a true stationary value. It has a stationary value with respect to x only. Hence a necessary and sufficient condition for the existence of a stationary value is that the first derivative vanish in *all* directions.

3. Extreme Values of an Integral.—In the problems of the calculus of variations the dependent variable is the value of the integral, and the independent variable is the form of the function that appears as an argument in the integrand. If a given form of the function $y = y(x)$ gives the integral a minimum value, any neighboring function must give the integral a value equal to or greater than the minimum. To make such a statement precise, it is necessary to define what is meant by *neighboring functions*. This can be done as follows: Let $y = f(x)$ be the function for which the integral has its extreme value. Let $\eta(x)$ be another function of x that is bounded, continuous, and has a continuous first derivative. Then it is convenient to define

$$y = f(x) + \alpha\eta(x) \tag{5-7}$$

as a family of functions in the neighborhood of $f(x)$. α is a parameter that may be given any positive or negative value, and the functions of the family may be taken as close as desired to $f(x)$ by giving to α a sufficiently small value.

If the integral has a maximum value when its values for only certain forms of the function $\eta(x)$ are considered, it is said to have a relative maximum. To have an absolute maximum it must have a value larger than or equal to the value given when *any* form of the function $\eta(x)$ is used.

If the integral I is evaluated for the family of functions represented by equation (5-7), I will be a function of α. If, then, the integral is to have an extreme value when $y = f(x)$, it is necessary that $(dI/d\alpha)_{\alpha=0} = 0$ and that this be true for all functions $\eta(x)$.

Problem 1. Consider the family of curves $y = x + \alpha \sin nx$, with n an integer, that lie in the neighborhood of the function $y = x$.

Show by actual evaluation that the integral of $(dy/dx)^2$ between $x = 0$ and $x = 2\pi$ is smaller for $y = x$ than for any other member of this family.

Problem 2. Consider the family of curves

$$y = Ax + B \sin nx + \alpha \sin (n + 2)x$$

that lie in the neighborhood of $y = Ax + B \sin nx$. Show that $I = \int_0^{2\pi} (dy/dx)^2 \, dx$ is less for $y = Ax + B \sin nx$ than for any other member of this family but that for the curves $y = Ax + (B - \alpha) \sin nx$ it can be still less.

Problem 3. Evaluate the integral $I = \int_0^X [(dy/dx)^2 + 4y]dx$ for $y = x + x^2 + \alpha[x^3 - (x^4/X)]$, and show that the minimum value occurs for $\alpha = 0$ and that $dI/d\alpha = 0$ when $\alpha = 0$.

It is customary in the calculus of variations to deal with variations of the integral and the quantities in the integrand rather than with their differential quotients. The variations are designated by the sign δ. With the family of functions defined by equation (5-7),

$$\delta y = \alpha\eta(x) \qquad \text{and} \qquad \delta y' = \alpha\eta'(x) = \frac{d}{dx}(\delta y) \tag{5-8}$$

The last of these equalities is true because

$$\frac{d}{dx}(y + \delta y) = y' + \delta y' = \frac{d}{dx}[y + \alpha\eta(x)] = y' + \alpha\eta'(x)$$

Each of these variations is a function not only of x but also of α and is in fact proportional to α. For the function

$$y = f(x) + \alpha\eta(x),$$

the integral will also have a value that depends on α, although it is not generally a linear function of α. Under suitable assumption as to the form of the integrand, the integral can be expanded as a function of α by differentiating with respect to the parameter α under the integral sign. This leads to

$$\int_{x_1}^{x_2} \Phi(y + \delta y, x, y' + \delta y')dx = \int_{x_1}^{x_2} \Phi(y,x,y')dx + \alpha \int_{x_1}^{x_2} \left[\frac{\partial\Phi}{\partial y}\eta(x) + \frac{\partial\Phi}{\partial y'}\eta'(x)\right] dx + \cdots \tag{5-9}$$

If the integral has an extreme value for $\alpha = 0$, the first derivative with respect to α will be zero and the second term in this expansion will vanish. The term containing α is called the *first variation* of the integral.

$$\delta I = \alpha \int_{x_1}^{x_2} \left[\frac{\partial \Phi}{\partial y} \eta(x) + \frac{\partial \Phi}{\partial y'} \eta'(x) \right] dx$$

A necessary condition for the existence of an extreme value is then that, for any value of α,

$$\delta I = 0 \tag{5-10}$$

This condition is a necessary condition for a maximum or a minimum. It is not the only necessary condition, nor is it at all a sufficient condition. It is, however, the condition that is of importance in the variation principles of mechanics. Such principles are concerned with the vanishing of the first variation rather than with the existence of a maximum or a minimum. For further discussion of the general variation problem, reference should be made to a treatise on the calculus of variations.

4. The Euler-Lagrange Equation.—The condition in equation (5-10) requires that

$$\int_{x_1}^{x_2} \left[\frac{\partial \Phi}{\partial y} \eta(x) + \frac{\partial \Phi}{\partial y'} \eta'(x) \right] dx = 0 \tag{5-11}$$

for all suitable functions $\eta(x)$. A function is *suitable* if it is continuous, has a continuous first derivative, and satisfies the boundary conditions. If the values of y_1 and y_2 are specified in the problem, $\eta(x_1)$ and $\eta(x_2)$ must both vanish, since there is no point to considering curves that do not connect the specified points. If the values of y_1 and y_2 are not specified, no such restriction can be imposed on $\eta(x)$, and equation (5-11) must hold whether $\eta(x)$ vanishes for $x = x_1$ and $x = x_2$ or not (see Fig. 5-1).

With these restrictions it is possible to carry out the second integration, by parts, and

$$\int_{x_1}^{x_2} \left[\frac{\partial \Phi}{\partial y} - \frac{d}{dx} \frac{\partial \Phi}{\partial y'} \right] \eta(x) dx + \frac{\partial \Phi}{\partial y'} \eta(x) \Big|_{x_1}^{x_2} = 0 \tag{5-12}$$

This condition can be satisfied for almost any $y(x)$ if the $\eta(x)$ is properly selected; but, to be sure that it is satisfied for any $\eta(x)$ that could be used, it is necessary that $(\partial\Phi/\partial y')\eta(x) = 0$ at x_1 and x_2 and also that

$$\frac{d}{dx}\frac{\partial\Phi}{\partial y'} - \frac{\partial\Phi}{\partial y} = 0 \tag{5-13}$$

for all values of x from x_1 to x_2. Equation (5-13) is the Euler-Lagrange equation of the variation problem. It is always a

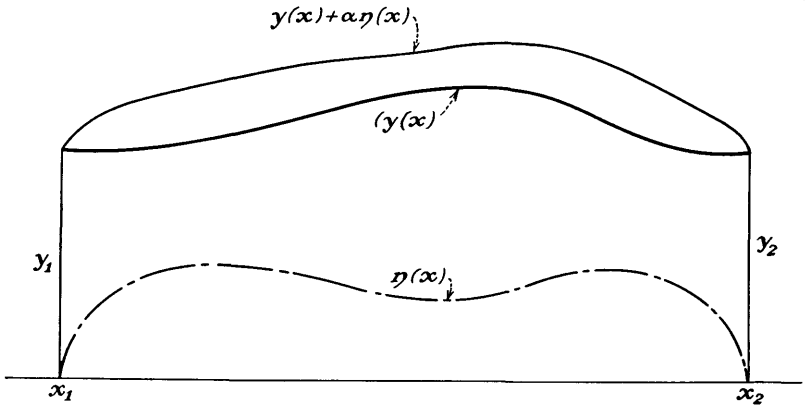


FIG. 5-1.—Sketch indicating the way $\eta(x)$ is added to $y(x)$ to give a function which satisfies the boundary conditions at $x = x_1$ and $x = x_2$, but which differs from $y(x)$ at other points.

necessary condition for the existence of an extreme value of the integral. If the values of y are specified at x_1 and x_2, a condition on $\partial\Phi/\partial y'$ at these points is not necessary, for all suitable $\eta(x)$'s will vanish there.

Problem 4. Show that, if $\Phi = (dy/dx)^2$, $y = Ax + B \sin nx$, and $\eta(x) = \sin (n + s)x$, $\delta I = 0$ even though this form of y does not give the integral of Φ from 0 to 2π a stationary value.

As an illustration of the use of the Euler-Lagrange equation consider the integral in equation (5-2). In the notation of equation (5-13)

$$\Phi = (1 + y'^2)^{\frac{1}{2}} \qquad \frac{\partial\Phi}{\partial y} = 0 \qquad \frac{\partial\Phi}{\partial y'} = \frac{y'}{(1 + y'^2)^{\frac{1}{2}}}$$

$$\frac{d}{dx}\frac{\partial\Phi}{\partial y'} = \frac{d^2y}{dx^2}\frac{1}{(1 + y'^2)^{\frac{3}{2}}} = 0$$

Thus it is necessary that the second derivative of y with respect to x shall be zero. This leads to a straight line whose equation contains two arbitrary constants which can be adjusted to make the line pass through the desired points.

Problem 5. Find the curve between the points ($x = 0$, $y = 0$) and ($x = x_1$, $y = y_1$) for which the integral of

$$\Phi = 2\left(\frac{dy}{dx}\right)^2 - \left(\frac{y}{x}\right)\left(\frac{dy}{dx}\right) + \frac{1}{2}\left(\frac{y}{x}\right)^2$$

has a stationary value.

Problem 6. Given two points and a line that are coplanar, find the form of the curve between the two points, and lying in the plane, that will generate a surface of minimum area when rotated about the straight line as an axis.

Problem 7. Find the curve connecting two given points down which a particle will slide in the shortest possible time.

Problem 8. Find the curve down which a particle will slide in the shortest possible time from a given point to a given vertical line.

5. Variation Problems with Several Dependent Variables.—The common case of variation principles in mechanics is that in which the integrand of the integral whose stationary value is sought contains one independent variable but more than one dependent variable. In this case it is necessary to admit arbitrary, independent variations of all the dependent variables. The first variation is then

$$\delta I = \int \left[\frac{\partial \Phi}{\partial y}\,\delta y + \frac{\partial \Phi}{\partial y'}\,\delta y' + \frac{\partial \Phi}{\partial z}\,\partial z + \frac{\partial \Phi}{\partial z'}\,\delta z' \cdots\right] dx$$

In the case of fixed limits, the partial integration gives

$$\delta I = \int_{x_1}^{x_2} \left[\left(\frac{\partial \Phi}{\partial y} - \frac{d}{dx}\frac{\partial \Phi}{\partial y'}\right)\delta y + \left(\frac{\partial \Phi}{\partial z} - \frac{d}{dx}\frac{\partial \Phi}{\partial z'}\right)\delta z + \cdots\right] dx \quad (5\text{-}14)$$

Since the variations of the different variables are independent, the vanishing of this expression requires the separate vanishing of the quantities in parentheses. This gives a set of simultane-

ous differential equations for the determination of the desired functions.

As an example, consider the problem of determining the shortest line between two points in three dimensions. In Cartesian coordinates the integral is

$$L \int_{x_1}^{x_2} (1 + y'^2 + z'^2)^{\frac{1}{2}}\, dx$$

The two simultaneous Euler-Lagrange equations are then

$$\frac{d}{dx}\frac{y'}{(1 + y'^2 + z'^2)^{\frac{1}{2}}} = 0 \qquad \frac{d}{dx}\frac{z'}{(1 + y'^2 + z'^2)^{\frac{1}{2}}} = 0$$

The solution of these gives y and z as functions of x and four arbitrary constants, which can be evaluated to make the curve pass through the desired points.

Problem 9. Find x and y as functions of t so that the integral

$$I = \int_{t_1}^{t_2} \left\{ \frac{m}{2} \left[\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 \right] - mgy \right\} dt$$

has a stationary value. Assume that x and y are given at t_1 and t_2.

Problem 10. Find the differential equations whose solution gives x and y as a function of the time such that the integral

$$I = \int_{t_1}^{t_2} \left\{ \frac{m}{2} \left[\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 \right] - a(x^2 + y^2) - bxy \right\} dt$$

has a stationary value. Take the end points as fixed.

6. Problems with Auxiliary Conditions.—In some cases it is desired to find the curve along which a given integral has an extreme value while at the same time certain relationships exist between the variables. For example, it might be desired to find the shortest line that connects two points and at the same time lies on a given surface. The problem of the shortest line between two points in three dimensions is one involving two dependent variables. If it is required that the line lie in the surface defined by $g(x,y,z) = 0$, a straightforward procedure is to use this relationship to express one of the variables, say y, in terms of the other two. The variation problem is then reduced to one of a single dependent variable,

To illustrate the procedure for the case of two dependent variables y and z and a single condition, let $\Phi(x,y,z,y',z')$ be the function whose integral between x_1 and x_2 is to be given an extreme value under the restriction that $g(x,y,z) = 0$. The first variation is

$$\delta I = \int_{x_1}^{x_2} \left[\left(\frac{\partial \Phi}{\partial y} - \frac{d}{dx} \frac{\partial \Phi}{\partial y'} \right) \delta y + \left(\frac{\partial \Phi}{\partial z} - \frac{d}{dz} \frac{\partial \Phi}{\partial z'} \right) \delta z \right] dx \quad (5\text{-}15)$$

This integral must vanish if the integral I has an extreme value. However, the vanishing of the integral does not require the independent vanishing of the two expressions in parentheses. δy and δz are not independent but are connected by the equation $g(x,y,z) = 0$. From this equation it follows that for any given value of x

$$\frac{\partial g}{\partial y} \delta y = - \frac{\partial g}{\partial z} \delta z \quad (5\text{-}16)$$

so that δz can be expressed as a product of δy and a function of x. With this value of δy the expression for the first variation becomes

$$\delta I = \int_{x_1}^{x_2} \left\{ \frac{\partial \Phi}{\partial y} - \frac{d}{dx} \frac{\partial \Phi}{\partial y'} - \left[\left(\frac{\partial \Phi}{\partial z} - \frac{d}{dx} \frac{\partial \Phi}{\partial z'} \right) \Big/ \frac{\partial g}{\partial z} \right] \frac{\partial g}{\partial y} \right\} \delta y \, dx$$

$$= \int_{x_1}^{x_2} \left[\frac{\partial \Phi}{\partial y} - \frac{d}{dx} \frac{\partial \Phi}{\partial y'} - \lambda(x) \frac{\partial g}{\partial y} \right] \delta y \, dx = 0 \quad (5\text{-}17)$$

Since δy is now entirely arbitrary and is not subject to any restrictions due to the auxiliary condition, equation (5-17) requires that

$$\frac{d}{dx} \frac{\partial \Phi}{\partial y'} - \frac{\partial \Phi}{\partial y} + \lambda(x) \frac{\partial g}{\partial y} = 0 \quad (5\text{-}18)$$

The function $\lambda(x)$ is also involved in the relationship

$$\frac{d}{dx} \frac{\partial \Phi}{\partial z'} - \frac{\partial \Phi}{\partial z} + \lambda(x) \frac{\partial g}{\partial z} = 0 \quad (5\text{-}18a)$$

Equations (5-18) and (5-18a) are those obtained by writing the Euler-Lagrange equations for a problem in which the integrand is $\Phi - \lambda(x)g(x,y,z)$ and no additional conditions need to be

considered. However, in this process an additional function $\lambda(x)$ has been introduced so that three functions $y(x)$, $z(x)$, and $\lambda(x)$ are to be determined by means of three equations, (5-18), (5-18a), and the auxiliary equation $g(x,y,z) = 0$.

The rule for treating a variation problem with algebraic auxiliary conditions is to form the function $F = \Phi + \Sigma\lambda_i g_i$ and to write the Euler-Lagrange equations for a stationary value of F. These will be of the form of (5-18) and (5-18a) and combined with the equation of condition will serve to determine the dependent functions and the auxiliary functions $\lambda(x)$.*

Problem 11. Find y and z as functions of x such that the first variation of $\int_{x_1}^{x_2} (y'^2 + z'^2 - ay)dx$ is zero and $z = by$. (b is a constant.)

Problem 12. Find y and z as functions of x such that the first variation of $\int_{x_1}^{x_2} (y'^2 + z'^2 - az^2)dx$ is zero and $z = b + cy$.

7. Isoperimetric Problems.—Sometimes the auxiliary condition in a variation problem is not a simple relationship between the variables, such as was treated in the previous section, but is the requirement that a certain integral shall have a predetermined value. Such problems are called *isoperimetric problems* because of their relationship to the classical problem of finding the closed curve of given perimeter that encloses the maximum area. The general problem can be formulated as follows: Find the form of y as a function of x such that the integral

$$I = \int_{x_1}^{x_2} \Phi(x,y,y')dx \tag{5-19}$$

has a stationary value with respect to those functions for which

$$S = \int_{x_1}^{x_2} \sigma(x,y,y')dx \tag{5-19a}$$

has the prescribed value S_0. We shall consider here only the case in which $y(x_1)$ and $y(x_2)$ are prescribed and not subject to variation. The generalization of the problem to include more than one dependent variable will also not be explicitly considered.

* The conditions under which a problem can be treated in this way are discussed by Forsyth, "Calculus of Variations," p. 433, Cambridge University Press, London.

As in the case of algebraic conditions, it is necessary to find a family of comparison curves that satisfy the auxiliary condition, *i.e.*, that for which S has the value S_0. Let $y(x)$ be the function that represents the desired solution of the problem, and let $\xi(x)$ and $\eta(x)$ be two different functions of x that vanish at x_1 and x_2. Then

$$y + \delta y = y + \alpha\xi + \beta\eta$$

represents a two-parameter family of curves that includes the curve y when $\alpha = \beta = 0$. The integral S along one of these curves is a function of α and β and can be expanded in terms of them.

$$S(\alpha,\beta) = S_0 + \alpha S_\alpha + \beta S_\beta + \cdots$$

where

$$S_0 = \int_{x_1}^{x_2} \sigma(x,y,y')dx \qquad S_\alpha = \int_{x_1}^{x_2} \left(\frac{\partial\sigma}{\partial y} - \frac{d}{dx}\frac{\partial\sigma}{\partial y'}\right)\xi(x)dx$$

$$S_\beta = \int_{x_1}^{x_2} \left(\frac{\partial\sigma}{\partial y} - \frac{d}{dx}\frac{\partial\sigma}{\partial y'}\right)\eta(x)dx$$

The requirement that $S(\alpha,\beta) = S_0$ constitutes a relationship between α and β that for small values of α and β can be approximated by $\beta = -(S_\alpha/S_\beta)\alpha$. Thus if $\delta y = \alpha[\xi - (S_\alpha/S_\beta)\eta]$, the family of curves $y + \delta y$ satisfies the auxiliary condition for small values of α.

If, with respect to this family of curves, the integral has an extreme value when $\alpha = 0$, it follows that

$$\delta I = \alpha \int_{x_1}^{x_2} \left(\frac{\partial\Phi}{\partial y} - \frac{d}{dx}\frac{\partial\Phi}{\partial y'}\right)\left(\xi - \frac{S_\alpha}{S_\beta}\eta\right)dx = 0 \qquad (5\text{-}20)$$

Let

$$\int_{x_1}^{x_2} \left(\frac{\partial\Phi}{\partial y} - \frac{d}{dx}\frac{\partial\Phi}{\partial y'}\right)\eta(x)dx = I_\beta$$

Equation (5-20) can then be written

$$\delta I = \alpha \int_{x_1}^{x_2} \left(\frac{\partial\Phi}{\partial y} - \frac{d}{dx}\frac{\partial\Phi}{\partial y'}\right)\xi(x)dx - \alpha\frac{I_\beta}{S_\beta}S_\alpha \qquad (5\text{-}21)$$

$$= \alpha \int_{x_1}^{x_2} \left[\frac{\partial\Phi}{\partial y} - \frac{d}{dx}\frac{\partial\Phi}{\partial y'} - \frac{I_\beta}{S_\beta}\left(\frac{\partial\sigma}{\partial y} - \frac{d}{dx}\frac{\partial\sigma}{\partial y'}\right)\right]\xi(x)dx = 0$$

The ratio I_β/S_β is fixed as soon as $\eta(x)$ is selected because of the definitions of I_β and S_β. But, in addition, equation (5-21) could be written

$$I_\alpha - \frac{I_\beta}{S_\beta} S_\alpha = 0 \tag{5-21a}$$

and since I_α and S_α can depend only on $\xi(x)$ while I_β and S_β depend on $\eta(x)$, it follows that I_β/S_β is a constant independent of $\eta(x)$. Equation (5-21) then also states merely that I_α/S_α is also a constant, independent of $\xi(x)$.

The formulation of equation (5-21) may be expressed in a simple rule. If the function $F = \Phi + \lambda\sigma$ is formed, a necessary condition for a stationary value of $I = \int_{x_1}^{x_2} \Phi \, dx$ with a prescribed value of $S = \int_{x_1}^{x_2} \sigma \, dx$ is

$$\frac{d}{dx}\frac{\partial F}{\partial y'} - \frac{\partial F}{\partial y} = 0 \tag{5-22}$$

together with the equation $S = S_0$. These two equations serve to determine $y(x)$ and the value of the constant λ.

To illustrate this method, consider the problem of finding a curve passing through the points $(x = 0, y = 0)$ and $(x = a, y = 0)$ for which the integral $I = \frac{1}{2}\int_0^a (dy/dx)^2 \, dx$ has a stationary value at the same time that $\int_0^a y^2 \, dx = S_0$.

Following the rule outlined above leads to the function $F = \frac{1}{2}(dy^2/dx) + \lambda y^2$, and equation (5-22) then takes the form

$$\frac{d^2y}{dx^2} - 2\lambda y = 0 \tag{5-23}$$

The general solution of this equation is

$$y = Ae^{\sqrt{2\lambda}x} + Be^{-\sqrt{2\lambda}x} \tag{5-23a}$$

The fact that the curve passes through the origin requires that $B = -A$, and the passing through $(x = a, y = 0)$ requires that

$$e^{\sqrt{2\lambda}a} - e^{-\sqrt{2\lambda}a} = 0$$

This fixes the value of $\sqrt{2\lambda}a$ as one of the series $n\pi i$ where n is an integer. Only the case with $n \neq 0$ leads to a useful solution. With the conditions at the end points the solution (5-23a) becomes

$$y = A_0 \sin \frac{n\pi x}{a} \tag{5-23b}$$

where A_0 is a new constant proportional to the original A. To evaluate A_0 use must be made of S_0.

$$A_0^2 \int_0^a \sin^2 \frac{n\pi x}{a}\, dx = S_0 = A_0^2 \frac{a}{2} \tag{5-23c}$$

This illustration is an example of a case in which there is an infinity of solutions giving stationary values.

It is to be noted that all the discussion in this chapter has referred to necessary conditions for a stationary value. If the solution in question represents a stationary value, the Euler-Lagrange equation will be satisfied. The question of the sufficiency of the conditions, of whether the satisfaction of the Euler-Lagrange equation guarantees a stationary value, is discussed in treatises on the calculus of variations.

Problem 13. The ends of a uniform string are fastened at given points. Find the equation of the curve in which it must hang in order that its center of mass be as low as possible.

Problem 14. A fence of length L is to be built so as to connect two points on a straight wall that are a distance $2a$ apart. Along what curve should the fence be built to enclose the maximum area? Consider only the case in which $L \leq \pi a$.

CHAPTER VI

HAMILTON'S PRINCIPLE

Hamilton's principle is a variation principle; it is equivalent to the Newtonian equations of motion and can be derived from them. Instead of describing the motion of a particle directly in terms of its acceleration, this principle describes the path in terms of a quantity whose integral along the path has a stationary value compared with other possible paths. From the statement that the variation of a certain integral is zero can be derived the differential equations that describe the motion. This principle provides a statement of the Newtonian equations that is independent of the coordinates used and from which the differential equations can be obtained with a minimum of effort. The variation principle is of little or no assistance in solving the equations, but it does provide a convenient means of writing them in any desired coordinates.

Hamilton's principle is not the only variation principle that can be used for mechanical problems. Others, notably the principle of least action, can be, and often have been, used. However, in this chapter only Hamilton's principle will be treated.

1. Derivation of Hamilton's Principle from Newton's Laws. Consider a mechanical system of n particles whose coordinates are $x_1, y_1, z_1, x_2, y_2, z_2, \ldots, x_n, y_n, z_n$. The motion of the system is known when the value of every coordinate is known as a function of the time. Suppose the system moves from a certain configuration given by $x_1', \ldots, z_n'$ at the time t' to another configuration given by $x_1'', y_1'', \ldots, z_n''$ at the time t''. During all of the motion between these two configurations the Newtonian equations of motion will be followed, and the acceleration of each particle will be given by the total force acting on it. This motion can be described by expressing each

coordinate as a function of the time. There are then $3n$ dependent variables depending on the one independent variable t. These functions can be written

$$x_1 = x_1(t), \qquad y_1 = y_1(t), \qquad \ldots, \qquad z_n = z_n(t) \quad (6\text{-}1)$$

Now consider some other way in which the system might have moved from the initial configuration to the final configuration in the same amount of time, $t'' - t'$. This new motion is to be one that satisfies the geometrical conditions, or the constraints, of the problem. It will not, however, satisfy the equations of motion. If this new motion is just slightly different from the original motion, the coordinates, as functions of the time, can be written

$$x_1(t) + \delta x_1(t), \qquad y_1(t) + \delta y_1(t), \qquad \ldots, \qquad z_n(t) + \delta z_n(t)$$

The variation of a coordinate x is a function of the time and is the difference between the x coordinate of the comparison path and that of the true path. It is to be assumed that the true path is a continuous function with continuous first derivatives, as it must be to satisfy Newton's equations. Similarly, the comparison paths must be functions of the time that are continuous and have continuous first derivatives. It is also specified that the true path and the comparison path lead from the same initial configuration to the same final configuration in the same time. On this account,

$$\delta x_1(t') = \delta x_1(t'') = \delta y_1(t') = \delta y_1(t'') = \cdots = \delta z_n(t'') = 0 \quad (6\text{-}2)$$

The true path was originally defined in terms of the Newtonian equations of motion. The object now is to translate this definition into a definition in terms of the properties of this path compared with these various other possible paths. For the true path there are the $3n$ equations

$$m_i \frac{d^2 x_i}{dt^2} = X_i$$

The quantity X_i may be a function of the coordinates, a function of the time explicitly, or both. It may be considered,

however, as a function of the time only, since the dependence on the coordinates is a dependence upon the positions of the particles and these are uniquely determined by the time along any path that may be considered. If, now, each component of the force X_i is multiplied by the variation of path in the direction of the force and all the resulting equations are added together, the result is

$$\begin{aligned}\delta U &= \sum_i (X_i\,\delta x_i + Y_i\,\delta y_i + Z_i\,\delta z_i) \\ &= \sum_i m_i \left(\frac{d^2x_i}{dt^2}\,\delta x_i + \frac{d^2y_i}{dt^2}\,\delta y_i + \frac{d^2z_i}{dt^2}\,\delta z_i\right) \qquad (6\text{-}3) \\ &= \sum m_i \left[\frac{d}{dt}(\dot{x}_i\,\delta x_i + \dot{y}_i\,\delta y_i + \dot{z}_i\,\delta z_i)\right. \\ &\qquad \left. - \dot{x}_i\,\delta\dot{x}_i - \dot{y}_i\,\delta\dot{y}_i - \dot{z}_i\,\delta\dot{z}_i\right]\end{aligned}$$

The quantity δU is defined by the first equality in equation (6-3). It is the work done by the forces of the system during the infinitesimal displacement $\delta x_i, \ldots, \delta z_n$ and is a function of the time and the independent coordinates of the system. If the forces do not depend explicitly on the time, δU can be expressed as a function of the coordinates only. The last part of (6-3) represents the variation of the kinetic energy δT. Hence the equation can be written

$$\delta T + \delta U = \sum m_i \frac{d}{dt}(\dot{x}_i\,\delta x_i + \dot{y}_i\,\delta y_i + \dot{z}_i\,\delta z_i) \qquad (6\text{-}4)$$

In all these expressions t is the independent variable. If both sides of (6-4) are integrated with respect to this independent variable between the limits t' and t'', the result is

$$\int_{t'}^{t''} (\delta T + \delta U)dt = \delta \int_{t'}^{t''} T\,dt + \int_{t'}^{t''} \delta U\,dt = 0 \qquad (6\text{-}5)$$

The integral of the right-hand side of (6-4) is zero because all of the variations are zero at both limits. Equation (6-5) is a property of the path that satisfies the equations of motion, and

this property furnishes a way of defining the true path of the system.

In the special case in which the forces are conservative, *i.e.*, when they can be derived from a potential energy, δU is the negative of the variation of the potential energy, so that

$$\delta \int_{t'}^{t''} (T - V)dt = \delta \int_{t'}^{t''} L\, dt = 0 \qquad (6\text{-}5a)$$

The quantity $T - V$ is denoted by L and is called the *Lagrangian function* or the kinetic potential of the system. The Lagrangian function can be expressed in any convenient coordinates, and the variation principle will still apply.

Hamilton's principle, then, states that for the motion of a mechanical system

$$\delta \int_{t'}^{t''} L(q_1, q_2, \ldots, q_n, \dot{q}_1, \dot{q}_2, \ldots, \dot{q}_n, t)dt = 0 \qquad (6\text{-}6)$$

In this equation the q's represent the coordinates necessary to specify the configuration of the system. The time appears explicitly in the Lagrangian function only in case the forces are explicit functions of the time, or the coordinates used are in motion. In the simple conservative cases the Lagrangian function depends upon the coordinates and their first derivatives only. If, as has been assumed, the coordinates are all independent, the treatment of the previous chapter shows that the path is described by the set of differential equations

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}_i} - \frac{\partial L}{\partial q_i} = 0 \qquad (6\text{-}7)$$

The coordinates in the Lagrangian function may be independent from the beginning, as in the case of the free motion of particles under their mutual forces. On the other hand, the particles may be constrained to move upon certain lines or surfaces. In this case the original $3n$ coordinates can be reduced in number, by the use of the equations that express the constraints, until a smaller number of independent coordinates is left. The use of equation (6-7) as it stands implies that this latter process has been carried out, if necessary.

The Euler-Lagrange equations for Hamilton's principle [equation (6-7)] are usually called simply Lagrange's equations. They contain nothing more than was contained in the original Newtonian equations, but they have the decided advantage that the coordinates may be of any kind whatever. It is necessary only to write the potential and the kinetic energies in the desired coordinates, to obtain the equations of motion by simple differentiation. This is usually much simpler than transforming the differential equations themselves.

Although Lagrange's equations have been obtained here from Hamilton's principle, they can also be obtained directly by transformation from Newton's equations. This serves to emphasize the fact that Lagrange's equations and Newton's equations are entirely equivalent, and the more convenient form should be used.

From the form of the Lagrangian function it is often possible to obtain one or more integrals of the motion. If a particular coordinate q_s is not contained in the Lagrangian function, equation (6-7) shows at once that

$$\frac{\partial L}{\partial \dot{q}_s} = \text{const.} \tag{6-8}$$

For this reason it is usually desirable to make such transformations of coordinates that as few as possible appear explicitly in the Lagrangian function. By this process the conservation of momentum and of angular momentum can be established for those systems in which they hold.

2. Illustration of Lagrange's Equation for Conservative Systems.—To illustrate the use of equation (6-7), consider a particle moving in a plane and attracted toward the origin of coordinates with a force inversely proportional to the square of the distance from it. With plane polar coordinates one has simply

$$V = -\frac{k}{r} \qquad \text{and} \qquad T = \frac{m}{2}(\dot{r}^2 + r^2\dot{\theta}^2)$$

From this

$$L = T - V = \frac{m}{2}(\dot{r}^2 + r^2\dot{\theta}^2) + \frac{k}{r}$$

and the derivatives are

$$\frac{\partial L}{\partial \dot{r}} = m\dot{r} \qquad \text{and} \qquad \frac{\partial L}{\partial r} = mr\dot{\theta}^2 - \frac{k}{r^2}$$

which give for this equation of motion

$$m\frac{d^2r}{dt^2} - mr\dot{\theta}^2 + \frac{k}{r^2} = 0$$

For the other equation in the variable θ,

$$\frac{\partial L}{\partial \dot{\theta}} = mr^2\dot{\theta} \qquad \text{and} \qquad \frac{\partial L}{\partial \theta} = 0$$

so that this equation of motion is

$$m\frac{d}{dt}(r^2\dot{\theta}) = 0$$

These are the equations of motion obtained in Chap. II by the much more laborious process of changing the variables. The last of these is an illustration of the use that can be made of the absence of a coordinate from the Lagrangian function. Since θ is not explicitly present in L, the derivative of L with respect to $\dot{\theta}$ is a constant.

Problem 1. A mass m moves on the x axis under the influence of the force $X = -m\omega^2 x$.

a. If the motion of the particle is given by $x = A \sin \omega t$, find the time integral of the Lagrangian function from $t = 0$ to $t = \pi/4\omega$.

b. Consider also the different motion given by the equation $x = A(\sin \omega t - \alpha \sin 4\omega t)$. This motion coincides with the previous motion at the times $t = 0$ and $t = \pi/4\omega$. Find the time integral of the Lagrangian function between the same limits for this second motion.

Problem 2. Find the differential equations of motion, in spherical polar coordinates, for a particle attracted toward the origin with a force that is a function of the distance only.

Problem 3. A particle moves on the surface of a sphere under the influence of gravity. Use spherical polar coordinates, and find the differential equations of motion.

Problem 4. Find the differential equations of motion for a free particle in cylindrical coordinates.

Problem 5. A particle slides along a wire bent into the shape of a circular helix whose equation is $z = a\theta$. The particle is also attracted toward a point on the z axis with a force proportional to the distance from it. Write and solve the equations of motion.

Problem 6. A mass m is lying on a smooth table and is attached by a cord of length l, which passes through a hole in the table, to a mass M suspended beneath. Write the differential equations of motion. If m is given an initial velocity perpendicular to the cord, find the minimum velocity that will keep M from descending. Consider M to move in a vertical line only.

3. Problems Involving Constraints.—As was stated above, the paths with which the true path is to be compared, in applying Hamilton's principle, are those which satisfy the constraints of the problem. In the preceding problems it has been assumed that either there are no constraints to begin with or the number of coordinates has been reduced until those remaining are all independent. In the case of the particle moving on the surface of the sphere, if only two angular coordinates are used, they are independent, and so the constraint is said to have been removed. If Cartesian coordinates are used for this problem, an additional condition must be introduced to confine the motion to the sphere. It is often inconvenient to reduce the number of coordinates, and a better way is to treat the equations connecting the coordinates as auxiliary conditions, by a method similar to that described in the previous chapter.

The relationships between the coordinates, which may include the time, may be written in the form

$$\phi_s(q_1, q_2, \ldots, q_n, t) = 0 \tag{6-9}$$

In the preceding chapter it was shown that such an auxiliary condition can be taken into account by including it in the integrand of the variation problem. Hence the variation principle can be written

$$\delta \int_{t'}^{t''} \left(L + \sum_s \lambda_s \phi_s\right) dt = 0 \tag{6-10}$$

The corresponding Lagrangian equations, since the ϕ's do not contain the velocities, are

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}_i} - \left(\frac{\partial L}{\partial q_i} + \sum \lambda_s(t)\frac{\partial \phi_s}{\partial q_i}\right) = 0 \qquad (6\text{-}11)$$

In dealing with equations of this kind, it is important to remember that λ is an unknown function of the time. In simple cases it may turn out to be a constant, but this cannot be assumed in solving the equations.

As a simple illustration of this method, consider a bead that can slide without friction on a straight wire. Let the wire make an angle with the horizontal whose tangent is a, and then take the x-y axes in the vertical plane containing the wire. Let the x axis be horizontal. Consider the force of gravity and an attraction toward the y axis equal to k times the distance from it.

The kinetic energy is $(m/2)(\dot{x}^2 + \dot{y}^2)$, the potential energy is $(k/2)x^2 + mgy$, and the equation of constraint is

$$y - ax - b = 0$$

The variation equation is then

$$\delta \int_{t'}^{t''} \left[\frac{m}{2}(\dot{x}^2 + \dot{y}^2) - \frac{k}{2}x^2 - mgy + \lambda(y - ax - b)\right] dt = 0$$

From this the two differential equations of motion are

$$m\ddot{x} + kx + a\lambda(t) = 0$$
$$m\ddot{y} + mg - \lambda(t) = 0$$

The function $\lambda(t)$ can be eliminated between these two equations to give

$$m(\ddot{x} + a\ddot{y}) + kx + amg = 0$$

The equation of constraint gives $\ddot{y} = a\ddot{x}$, and therefore an equation in x alone can be written

$$m(1 + a^2)\ddot{x} + kx + amg = 0$$

The solution of this differential equation is immediately obvious as

$$x = -\frac{amg}{k} + A \sin\left[\left(\frac{k/m}{1 + a^2}\right)^{\frac{1}{2}} t - \epsilon\right]$$

From the equation of constraint, y is just $ax + b$, and $\lambda(t)$ can be determined from either differential equation.

$$\lambda(t) = +mg - \frac{kaA}{1 + a^2} \sin \left[\left(\frac{k/m}{1 + a^2} \right)^{\frac{1}{2}} t - \epsilon \right]$$

Problem 7. A mass m is fastened to one end of a string that passes over a fixed pulley. From the other end of the string is suspended a second pulley, and over it passes a string supporting masses m_1 and m_2. Use Lagrange's equation involving constraints to find the motion of the system.

Problem 8. Solve Prob. 5 by the use of Lagrange's equation with constraints.

Problem 9. Show, by carrying out the substitutions, that equation (6-11) can be obtained by eliminating some of the independent variables in Hamilton's principle through the equations of constraint.

Problem 10. A smooth wire is bent into the form of an inverted cycloid. Find the motion of a particle sliding on this wire under the influence of gravity.

4. Problems with Nonconservative Forces.—In the problems thus far treated it has been possible to express the forces in terms of the derivatives of a potential energy. This is not always possible, and systems in which it is not are called *nonconservative systems.* When a system is nonconservative, it is frequently because not all the bodies that act on each other are included as belonging to the system under consideration. In such cases, the forces due to the neglected bodies must be known as functions of the time and the positions and velocities of the particles that are included.

It is possible to use Lagrange's equations in arbitrary coordinates as equations of motion for nonconservative systems by including a generalized force term on the right-hand side. If the coordinates to be used are $q_1, q_2, \ldots, q_n$, the work done by the forces of the system when these coordinates are changed by small amounts can be written

$$\delta U = Q_1 \, \delta q_1 + Q_2 \, \delta q_2 + \cdots Q_n \, \delta q_n \qquad (6\text{-}12)$$

δU has the dimensions of work; but since the q's may not all have the dimensions of length, the Q_i's will not necessarily

have the dimensions of force. With the Q_i's as defined in equation (6-12), the Lagrangian differential equations of motion are

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}_i} - \frac{\partial L}{\partial q_i} = Q_i \tag{6-13}$$

In this equation, the function L contains the kinetic energy and the potential energy due to any conservative forces that may be acting. The remaining, nonconservative forces are described by the quantities Q_i.

Consider a particle constrained to move in a plane under the influence of an attraction toward the origin proportional to the distance from it and also of a force perpendicular to the radius vector. This latter force is inversely proportional to the distance of the particle from the origin and has the counterclockwise sense. The Lagrangian function is

$$L = \frac{m}{2}(\dot{r}^2 + r^2\dot{\theta}^2) - \frac{a}{2}r^2 \tag{6-14}$$

when plane polar coordinates are used. The nonconservative force does work only when the angle θ is changed and

$$\delta U = \frac{b}{r} r\,\delta\theta = b\,\delta\theta \tag{6-14a}$$

where b is the constant of proportionality. The equations of motion are then

$$\begin{aligned} m\ddot{r} - mr\dot{\theta}^2 + ar &= 0 \\ m\frac{d}{dt}(r^2\dot{\theta}) &= b \end{aligned} \tag{6-14b}$$

This is an illustration of the way in which the variation δU can be defined without the necessity of having a single-valued function U.

Problem 11. Write the differential equations of motion for a particle moving around an attracting center and opposed by a resistance proportional to the velocity.

5. Hamilton's Canonical Equations of Motion.—For many problems the methods just described serve to put the differential

equations of motion into the best form for integration. For some cases, however, and especially for general considerations, it is more convenient to use a system of $6n$ partial differential equations of the first order, instead of the $3n$ equations of the second order. These can be obtained as follows:

Let

$$p_i = \frac{\partial L}{\partial \dot{q}_i} \tag{6-15}$$

Then let

$$H = \Sigma p_i \dot{q}_i - L = H(p,q,t) \tag{6-16}$$

Since L is a homogeneous quadratic function of the $\dot{q}_i$'s, equations (6-15) are linear in the $\dot{q}_i$'s and can be solved for them in terms of the p_i's. With these solutions, the function H can be expressed as a function of the p_i's and q_i's. The differential of H can then be written in terms of either set of variables:

$$\begin{aligned} dH &= \sum p_i \, d\dot{q}_i + \sum \dot{q}_i \, dp_i - \sum \frac{\partial L}{\partial \dot{q}_i} \, d\dot{q}_i - \sum \frac{\partial L}{\partial q_i} \, dq_i - \frac{\partial L}{\partial t} \, dt \\ &= \sum \frac{\partial H}{\partial q_i} \, dq_i + \sum \frac{\partial H}{\partial p_i} \, dp_i + \frac{\partial H}{\partial t} \, dt \end{aligned} \tag{6-17}$$

Since the two forms of H are equal to each other for all values of the variables, the coefficients of the corresponding differentials must be equal. The coefficients of $d\dot{q}_i$ add to give zero because of the definition of the p_i's. There result, then, the equations

$$\dot{q}_i = \frac{\partial H}{\partial p_i} \qquad \frac{\partial L}{\partial q_i} = -\frac{\partial H}{\partial q_i} \qquad \frac{\partial H}{\partial t} = -\frac{\partial L}{\partial t}$$

By Lagrange's equations, $\dot{p}_i = \partial L / \partial q_i$, and therefore the equations are

$$\dot{q}_i = \frac{\partial H}{\partial p_i} \qquad \dot{p}_i = -\frac{\partial H}{\partial q_i} \tag{6-18}$$

The advantage that the Hamiltonian equations (6-18) have over the Lagrangian equations is that they contain $6n$ independent variables. All these can be transformed in the effort to get the equations in a form suitable for solution. Transforma-

tions of these variables that leave the form of the Hamiltonian function unchanged are called *canonical transformations.*

If a canonical transformation can be found such that the transformed Hamiltonian function is independent of one or more of the new coordinates, an integral of the equations is immediately obvious. The momenta conjugate to the missing coordinates are constant. The general procedure in solving the Hamiltonian equations is to find such a transformation that the new Hamiltonian function depends on momenta only. A general method for doing this is provided by the Hamilton-Jacobi partial differential equation, but the discussion of this equation is too extended to be given here.

Problem 12. Show that, if the Lagrangian function does not contain the time explicitly, the Hamiltonian function is equal to the total energy of the system.

Problem 13. Show that, if the Hamiltonian function does not contain the time explicitly, it is constant.

Problem 14. Write the Hamiltonian function and the equations of motion for a simple-harmonic oscillator. Then transform to the new variables α and φ' by the equations

$$x = \left(\frac{2\alpha}{m\omega}\right)^{\frac{1}{2}} \sin \varphi \qquad \text{and} \qquad p = (2m\alpha\omega)^{\frac{1}{2}} \cos \varphi \qquad (6\text{-}19)$$

Demonstrate that this is a canonical transformation by showing that the form of the equations of motion is unchanged. Solve the equation in these variables.

6. The Pendulum.—In Chap. II the simple pendulum was treated with the limitation that only very small swings were considered. With that restriction it was shown to have a sinusoidal motion with a frequency independent of the amplitude. However, it is also possible to treat the pendulum without this restriction, and the treatment provides an illustration of some of the methods used in more complicated cases.

Consider a pendulum composed of a mass m attached to the end of a light rod of length a. The other end is pivoted so that the mass can move freely in a vertical plane, in circles around the pivot.

a. The Differential Equations of Motion.—The differential equations of motion can be written in various ways. The most obvious way is to apply Newton's equations directly. Let the origin of the coordinates be taken at the pivoted end of the rod with the x axis horizontal and the y axis vertical. The forces acting on the mass m will be the tension in the rod directed along the rod and the force of gravity directed vertically downward. Taking components of these forces along the x and y axes, the equations of motion are

$$\left.\begin{aligned} m\ddot{x} &= -\frac{K}{a}x \\ m\ddot{y} &= -\frac{K}{a}y - mg \end{aligned}\right\} \tag{6-20}$$

These are relatively simple to write down, but they contain the tension K, which is an unknown function of the time and must be so treated in solving the equations. The approximate treatment, valid for small swings, consists in assuming that y is constant so that $\ddot{y} = 0$. This gives a constant value for K from the second equation, and the first one can be solved with this constant K.

This is also a case in which Lagrange's equation with constraints can be used.

$$T = \frac{m}{2}(\dot{x}^2 + \dot{y}^2) \qquad V = mgy \qquad x^2 + y^2 - a^2 = 0$$

From equation (6-11), it then follows that

$$\left.\begin{aligned} m\ddot{x} - 2\lambda x &= 0 \\ m\ddot{y} - 2\lambda y + mg &= 0 \end{aligned}\right\} \tag{6-20a}$$

In these equations the function 2λ replaces $-K/a$ of equations 6-20. This illustrates a general principle that the undetermined functions λ of equation (6-11) represent forces exerted by the constraints.

The most effective use of Lagrange's equation comes when a transformation of coordinates is made in the Lagrangian function. For the problem of the pendulum take polar coordinates

(r,θ) with the negative y axis corresponding to $\theta = 0$. The kinetic and potential energies can be written directly in these coordinates or transformed from the above expressions.

$$L = \frac{m}{2}(\dot{r}^2 + r^2\dot{\theta}^2) + mgr\cos\theta$$
$$\phi = r^2 - a^2 = 0$$

If account is taken of the constraint, the Lagrangian function becomes

$$L = \frac{m}{2}a^2\dot{\theta}^2 + mga\cos\theta$$

and there is only a single differential equation of motion

$$\ddot{\theta} + \frac{g}{a}\sin\theta = 0 \tag{6-20b}$$

If, instead, Lagrange's equation is used to take account of the constraint, there is an additional equation

$$\ddot{r} - r\dot{\theta}^2 - g\cos\theta - 2\frac{\lambda}{m}r = 0$$

for which λ may be determined as a function of the time. Since $r = a$, this leads to

$$\lambda = -\frac{m}{2}\dot{\theta}^2 - \frac{mg}{2a}\cos\theta$$

b. Solution of the Equation of Motion.—It is convenient to write equation (6-20*b*) in the form

$$\ddot{\theta} + 2\frac{g}{a}\sin\frac{\theta}{2}\cos\frac{\theta}{2} = 0 \tag{6-21}$$

and this will be taken as the equation of motion.

The Hamiltonian momentum is

$$p = ma^2\dot{\theta} \tag{6-22}$$

and the Hamiltonian function is

$$H = \frac{p^2}{2ma^2} + 2mga\sin^2\frac{\theta}{2} - mga \tag{6-23}$$

Since this does not contain the time, it is a constant, in this case the energy, and may be set equal to $C - mga$. Then it follows that

$$p = 2ma^2\left(\frac{g}{a}\right)^{\frac{1}{2}}\left(\frac{C}{2mga} - \sin^2\frac{\theta}{2}\right)^{\frac{1}{2}} \tag{6-24}$$

Some of the properties of the motion can be obtained directly from equation (6-24). Since p must be real, only those values

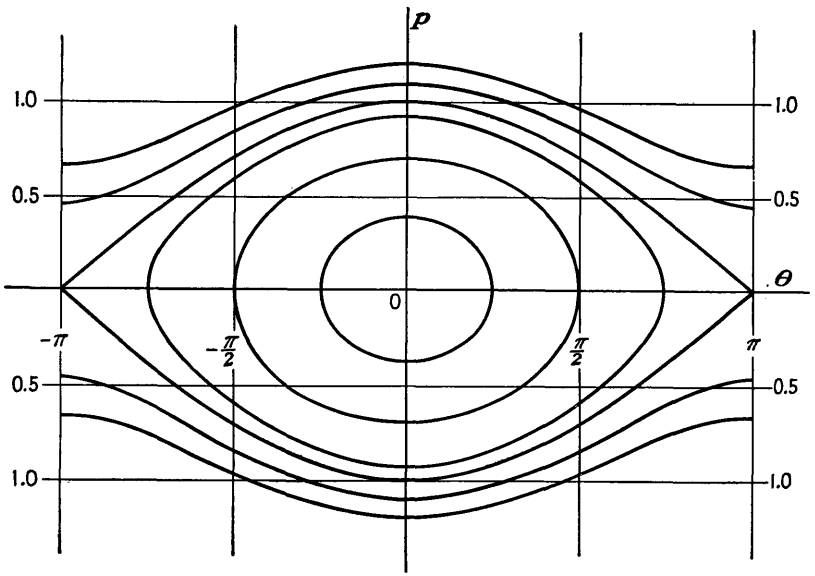


FIG. 6-1.—Some phase orbits of a pendulum.

of θ are permitted for which $\sin^2 \theta/2 < C/2mga$. The limiting case defines the maximum swing, so that the amplitude α is given by

$$\sin\frac{\alpha}{2} = \left(\frac{C}{2mga}\right)^{\frac{1}{2}} \tag{6-25}$$

If $C < 2mga$, there exists this limiting angle α, and θ oscillates between α and $-\alpha$. In the other case ($C > 2mga$) the pendulum turns clear over. Figure 6-1 shows p as a function of θ for various values of $C/2mga$. The lines in this figure may be regarded as orbits described by a point representing the position and momentum of the pendulum. The point describ-

ing an orbit of this kind does not move with constant velocity in this p-θ plane. Its velocity must be obtained by further integration of equation (6-24).

To find the period of the motion, p in equation (6-24) is replaced by its value in terms of $\dot{\theta}$. Then

$$2\left(\frac{g}{a}\right)^{\frac{1}{2}} dt = \frac{d\theta}{(\sin^2 \alpha/2 - \sin^2 \theta/2)^{\frac{1}{2}}} \tag{6-26}$$

To integrate this for the case of oscillatory motion ($\sin^2 \alpha/2 < 1$) let

$$\sin \varphi = \frac{\sin \theta/2}{\sin \alpha/2} \tag{6-26a}$$

Then

$$\left(\frac{g}{a}\right)^{\frac{1}{2}} (t - t_0) = \int_0^{\varphi_1} \frac{d\varphi}{(1 - \sin^2 \alpha/2 \sin^2 \varphi)^{\frac{1}{2}}} = F\left(\frac{\alpha}{2}, \varphi_1\right) \tag{6-26b}$$

This integral gives the time during which the angle φ increases from 0 to φ_1 and θ increases from 0 to the corresponding value of θ as given by equation (6-26a). When the integral is carried out to $\varphi_1 = \pi/2$, $\theta = \alpha$ and the time is a quarter period of oscillation. Hence, if T is the full period,

$$T = 4\left(\frac{a}{g}\right)^{\frac{1}{2}} \int_0^{\pi/2} \frac{d\varphi}{(1 - \sin^2 \alpha/2 \sin^2 \varphi)^{\frac{1}{2}}}$$

$$= 4\left(\frac{a}{g}\right)^{\frac{1}{2}} F\left(\frac{\alpha}{2}, \frac{\pi}{2}\right) \tag{6-27}$$

$F[(\alpha/2), \varphi]$ is an elliptic integral of the first kind whose values can be found in tables. For $\alpha/2 \to \pi/2$ the period T approaches infinity. This corresponds to the situation shown in Fig. 6-1 for $C/2mga = 1$. The orbit crosses the axis as a straight line so that the representative point never actually reaches the singular point $\theta = \pi$, $p = 0$.

If $C > 2mga$, the mass has enough kinetic energy at the bottom of its swing to carry it clear around. In this case equation (6-26b) becomes

$$t - t_0 = \left(\frac{2m}{C}\right)^{\frac{1}{2}} a \int_0^{\theta} \frac{d\theta/2}{[1 - (2mga/C) \sin^2 \theta/2]^{\frac{1}{2}}} \tag{6-28}$$

This again is an elliptic integral of the first kind that can be looked up in the tables. When $\theta = \pi$, the time is that necessary for a half revolution, so that if T is the time for a full revolution

$$T = 2a\left(\frac{2m}{C}\right)^{\frac{1}{2}} \int_0^{\pi} \frac{d\theta/2}{[1 - (2mga/C)\sin^2 \theta/2]^{\frac{1}{2}}} = 2a\left(\frac{2m}{C}\right)^{\frac{1}{2}} F\left(\beta, \frac{\pi}{2}\right) \quad (6\text{-}29)$$

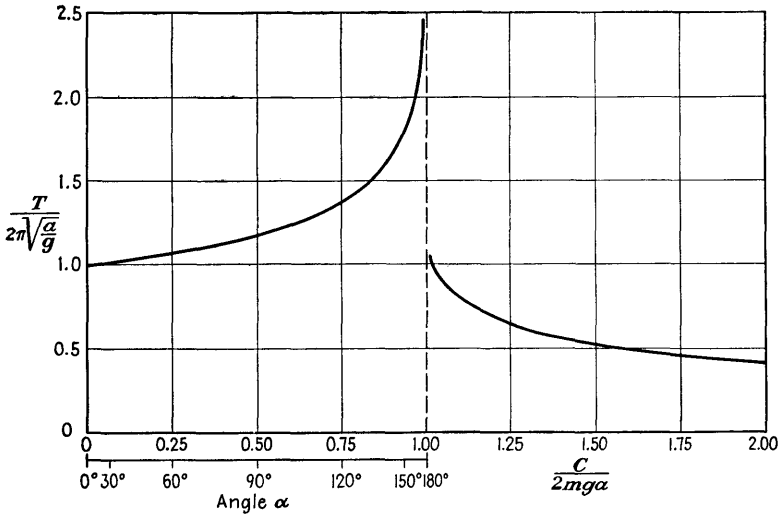


FIG. 6-2.—Period of a pendulum as a function of its amplitude. For a pendulum that makes a complete rotation, the period is expressed as a function of the energy.

where $\sin^2 \beta = 2mga/C$. The value of T as a function of $C/2mga$ is shown in Fig. 6-2. For the oscillatory case the magnitude of the amplitude is indicated also.

Problem 15. Obtain an approximate expression for the period of a pendulum with finite amplitude by expanding the integrand in equation (6-27) in powers of $\sin^2 \alpha/2$.

Problem 16. Show that for small values of the amplitude the orbit of the representative point in the p-θ plane is an ellipse and that for the limiting case of $\alpha = \pi$ the orbit is such that p is proportional to $\cos \theta/2$.

Problem 17. Show that in the limiting case of $C/2mga \to \infty$ the period given by equation (6-29) is just that to be expected when

gravity is neglected and the pendulum turns uniformly with its initial angular velocity.

References

BYERLY, W. E.: "Generalized Coordinates," Ginn & Company, Boston.

KÁRMÁN and BIOT: "Mathematical Methods in Engineering," McGraw-Hill Book Company, Inc., New York. (This contains a good account of the pendulum and of elliptic integrals.)

PAGE, L.: "Introduction to Theoretical Physics," D. Van Nostrand Company, Inc., New York.

WEBSTER, A. G.: "Dynamics of Particles," B. G. Teubner, Leipzig.

CHAPTER VII

THEORY OF VIBRATING SYSTEMS

In Chapter IV the vibrations of some systems of two and three degrees of freedom were studied and were expressed in terms of normal coordinates. In this chapter a similar treatment will be given to some systems of many degrees of freedom and to continuous strings. These last may be considered as vibrating systems with an infinite number of degrees of freedom.

1. General Theory of Normal Coordinates.—Because Lagrange's equation can always be used as the form of the equations of motion, the problem of treating a general vibrating system starts with the problem of writing the kinetic and potential energies. A vibrating system will have some position of equilibrium, and it is usually convenient to take the coordinates in such a way that they are all zero when the system is in its equilibrium configuration. The potential energy can then be expressed as a Taylor's series, which will be a power series in these coordinates. The constant term in this series can be set equal to zero by so defining the energy that it is zero in this equilibrium position. The coefficients of the first powers of the coordinates will be the first derivatives of the potential energy with respect to the coordinates. These will be zero, because the configuration is one of equilibrium. The first coefficients, then, different from zero, are those of the quadratic terms, and they will all be positive if the equilibrium is stable. If all the forces are harmonic, there will be no higher derivatives; and, in general, if only small displacements are considered, the higher terms can be neglected. Hence the potential energy will be a quadratic expression in the coordinates that, if the equilibrium is stable, will be a positive definite expression, *i.e.*, it will never be negative and will be zero only when all the coordinates are zero. It will also be possible to write the

kinetic energy as a positive definite quadratic expression in the time derivatives of the coordinates.

It is shown in works on algebra that it is always possible to reduce any two positive definite quadratic expressions to a sum of squares of quantities which are linear combinations of the original variables. Hence, it will always be possible to find such linear combinations of the originally selected coordinates that the kinetic and potential energies can be written as sums of squares of these combinations and their time derivatives. The linear combinations for which this is true are the normal coordinates of the problem. Thus it is always possible to find normal coordinates for any system whose potential energy can be expressed as a homogeneous quadratic function of the coordinates.

Problem 1. Show that, if the potential energy of a system is a sum of squares of the coordinates multiplied by constant coefficients and if the kinetic energy is a similar sum of the squares of the time derivatives of the coordinates, the differential equations of motion are those for simple-harmonic motion.

Problem 2. Write the exact expression for the potential energy of the system of two coupled pendulums connected by a spring, and show how it satisfies the various conditions presented in the above discussion of vibrating systems.

The methods of reducing the expressions for the potential and kinetic energies to sums of squares are in essence just the methods of Chap. IV.

2. Vibrations of a Loaded String.—In many cases the actual determination of the normal coordinates is a difficult matter because of the necessity of solving an algebraic equation of high order to find the normal frequencies. One case, however, which can be treated is that of a number of similar particles uniformly distributed along a string. This problem was first treated by Lagrange and is of especial interest because of the light it throws upon the vibrations of a continuous string.

The system consists of n particles, each of mass m, uniformly distributed along a string of length $(n + 1)a$. Let T be the tension in the string. This tension is assumed to be

the same at all points of the string and to be unaffected by the small displacements of the particles. The string is to be fastened at the end points and to have only a negligible mass. Let y_r be the displacement of the rth particle in a direction perpendicular to the string. All the displacements are to be taken in one plane. The differential equations of motion can be written down directly by considering the component of the

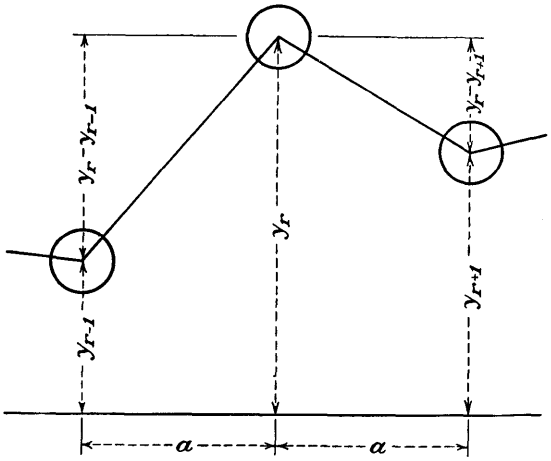


FIG. 7-1.—Specification of the coordinates used in describing the positions of the masses on a loaded string.

tension directed downward as shown in Fig. 7-1. The rth mass is pulled toward the axis by the force $(y_r - y_{r-1})T/a$ due to the tension on the left and by the force $(y_r - y_{r+1})T/a$ due to the tension on the other side. Hence

$$\frac{d^2y_r}{dt^2} = \frac{T}{ma}(y_{r-1} - 2y_r + y_{r+1}) \tag{7-1}$$

This equation can also be written down from the kinetic and potential energies.

Problem 3. Write the Lagrangian function for a stretched string, and show that it leads to equation (7-1). The potential energy can be taken as zero when all the particles are in a straight line and to be due to the work done in stretching the string against its tension.

The set of differential equations can be treated as in Chap. IV by substituting $y_r = A_r e^{i\omega t}$. For the first equation this leads to

$$\left(2 - \frac{ma\omega^2}{T}\right) A_1 - A_2 = 0 \qquad (7\text{-}2a)$$

since y_0 represents the displacement of the fixed end of the string and is zero. The next equation is

$$-A_1 + \left(2 - \frac{ma\omega^2}{T}\right) A_2 - A_3 = 0 \qquad (7\text{-}2b)$$

and the last equation will be

$$-A_{n-1} + \left(2 - \frac{ma\omega^2}{T}\right) A_n = 0 \qquad (7\text{-}2c)$$

since again, y_{n+1} must be taken as zero. The equation for the normal frequencies is then

$$\begin{vmatrix} C & -1 & 0 & 0 & 0 & \cdot & \cdot \\ -1 & C & -1 & 0 & 0 & \cdot & \cdot \\ 0 & -1 & C & -1 & 0 & \cdot & \cdot \\ 0 & 0 & -1 & C & -1 & \cdot & \cdot \\ 0 & 0 & 0 & -1 & C & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{vmatrix} = 0 \qquad (7\text{-}3)$$

(n rows and columns)

where

$$C = 2 - \frac{ma\omega^2}{T}$$

For $n = 1$ the equation is simply $C = 0$. This gives the frequency of a single particle at the center of a stretched string.

$$\omega^2 = \frac{2T}{ma} \qquad (7\text{-}4a)$$

For $n = 2$, $C^2 - 1 = 0$, and the values of ω^2 are

$$\left.\begin{aligned} \omega_1{}^2 &= \frac{T}{ma} \\ \omega_2{}^2 &= \frac{3T}{ma} \end{aligned}\right\} \qquad (7\text{-}4b)$$

For $n = 3$, $C(C^2 - 2) = 0$ so that

$$\left.\begin{aligned} \omega_1^2 &= (2 - \sqrt{2})\frac{T}{ma} \\ \omega_2^2 &= \frac{2T}{ma} \\ \omega_3^2 &= (2 + \sqrt{2})\frac{T}{ma} \end{aligned}\right\} \tag{7-4c}$$

These solutions for small values of n are obtained by simply expanding the determinant. In general, equation (7-3) is of the nth degree in C, but because of its symmetrical form it can be solved by a suitable change in variable.

For any value of C the determinant has a value that depends upon C and upon the number of rows and columns in the determinant. There is nothing else upon which this value can depend. Let the value of the determinant with n rows and columns be D_n, and let D_{n-1}, D_{n-2}, etc., be the values of the corresponding determinants with the smaller numbers of rows and columns. These quantities, D_n, D_{n-1}, etc., are all functions of C. Then, by expanding the determinant D_n in terms of the elements of its first row, there can be obtained a relationship between the determinants of the different orders.

$$D_n = CD_{n-1} - D_{n-2} \tag{7-5}$$

While evaluating the determinant, C is considered as a constant. This equation connects the values of determinants of different orders for the same value of C. The definition of C shows that it must always be less than 2, and consequently a possible substitution may be $C = 2 \cos \Theta$, where Θ is a new quantity defined by this relationship with C. If it can be shown in the end that the n different roots of equation (7-3) lead to values of C greater than -2, the substitution will be justified. With this substitution equation (7-5) becomes

$$D_n = 2 \cos \Theta D_{n-1} - D_{n-2} \tag{7-6}$$

This is a difference equation for D as a function of n, and its solution is

$$D_n = D \sin (n + 1)\Theta \tag{7-7}$$

The solution of a difference equation is similar to that of a differential equation in that it contains arbitrary constants. The quantity D in equation (7-7) is an arbitrary constant that can be evaluated from the known value of D_1. Since

$$D_1 = D \sin 2\Theta = C = 2 \cos \Theta$$

it follows that

$$D = \frac{1}{\sin \Theta} \tag{7-8}$$

and

$$D_n = \frac{\sin (n + 1)\Theta}{\sin \Theta} \tag{7-9}$$

Equation (7-9) gives the value of the determinant in terms of the quantity Θ, which is a known function of C, and hence the value of the determinant is given in terms of C.

From equation (7-9) it is evident that the values of Θ for which the determinant is zero are

$$\Theta_s = \frac{s\pi}{n + 1} \tag{7-10}$$

and hence the corresponding values of C are

$$C_s = 2 \cos \frac{s\pi}{n + 1} \tag{7-11}$$

In equations (7-10) and (7-11) s is an integer that can take any value from 1 to n. There are thus n different values of C and correspondingly n different frequencies with which the system can vibrate. These different frequencies will be designated by the subscript s.

$$\omega_s^2 = \frac{2T}{ma}\left(1 - \cos \frac{s\pi}{n + 1}\right) \tag{7-12}$$

Since equation (7-3) has just n roots, it is clear that all the roots have been found and it is not necessary to investigate further the generality of the substitution for C or of (7-7) as a solution of (7-6).

Problem 4. Show by substitution that equation (7-7) is a solution of (7-6) for any value of D.

Problem 5. Show that the normal frequencies for the cases of one, two, and three particles are given by equation (7-12).

After the frequencies are known, the next step is to solve the equations for the relative amplitudes A_r. The first two and the last of these equations are given by (7-2*a*), (7-2*b*), and (7-2*c*). The general equation is

$$-A_{r-1} + CA_r - A_{r+1} = 0 \tag{7-13}$$

This has just the form of equation (7-6) and therefore a similar solution. The difference lies in the boundary conditions, which require that $A_0 = A_{n+1} = 0$. If $A_r^{(s)}$ is the amplitude of the rth particle when $C = C_s$,

$$A_r^{(s)} = A^{(s)} \sin \frac{rs\pi}{n+1} \qquad \text{and} \qquad B_r^{(s)} = B^{(s)} \sin \frac{rs\pi}{n+1} \tag{7-14}$$

The quantities $B_r^{(s)}$ are the coefficients used with the negative ω_s in the exponent. The $2n$ arbitrary constants $A^{(s)}$ and $B^{(s)}$ provide the necessary number of arbitrary constants for a general solution.

Problem 6. Write the general solution of the problem of three equal masses on a stretched string.

Problem 7. In the case of three particles, evaluate the arbitrary constants in terms of the initial conditions when all the particles start from an initial displacement with zero initial velocity.

3. Normal Coordinates of a Loaded String.—With the values of the coefficients given in equations (7-14) the general solution of the set of differential equations can be written

$$\begin{aligned} y_r &= \sum_{s=1}^{n} A^{(s)} \sin \frac{rs\pi}{n+1} e^{i\omega_s t} + \sum_{s=1}^{n} B^{(s)} \sin \frac{rs\pi}{n+1} e^{-i\omega_s t} \\ &= \sum_{s=1}^{n} \sin \frac{rs\pi}{n+1} [A^{(s)} e^{i\omega_s t} + B^{(s)} e^{-i\omega_s t}] \\ &= \sum_{s=1}^{n} \sin \frac{rs\pi}{n+1} Y_s(t) \end{aligned} \tag{7-15}$$

The quantities $Y_s(t)$ are functions of t only and are the normal coordinates of the system. From their dependence on time it appears that they satisfy the differential equation

$$\frac{d^2Y_s}{dt^2} + \omega_s^2 Y_s = 0 \tag{7-16}$$

To express the kinetic energy in terms of these normal coordinates, equation (7-15) must be differentiated, squared, and summed over all values of r.

$$\dot{y}_r^2 = \sum_{s=1}^{n}\sum_{\sigma=1}^{n} \sin\frac{rs\pi}{n+1}\sin\frac{r\sigma\pi}{n+1}\dot{Y}_s\dot{Y}_\sigma \tag{7-17}$$

Then the kinetic energy

$$T = \frac{m}{2}\sum_{r=1}^{n}\dot{y}_r^2 = \frac{m}{2}\sum_{r=1}^{n}\sum_{s=1}^{n}\sum_{\sigma=1}^{n}\sin\frac{rs\pi}{n+1}\sin\frac{r\sigma\pi}{n+1}\dot{Y}_s\dot{Y}_\sigma$$

$$= \frac{m}{2}\frac{n+1}{2}\sum_{s=1}^{n}\dot{Y}_s^2 \tag{7-18}$$

The sum over r in this equation gives zero except when $\sigma = s$ so that only squares of the normal coordinates are left. In a similar manner it can be shown that the potential energy

$$V = \frac{m(n+1)}{4}\sum_{s=1}^{n}\omega_s^2 Y_s^2 \tag{7-19}$$

as is required of normal coordinates.

To express the normal coordinates in terms of the y_r's, multiply each y_r by $\sin r\sigma\pi/(n+1)$, and add them all together. Then from equation (7-15)

$$\sum_{r=1}^{n} y_r \sin\frac{r\sigma\pi}{n+1} = \sum_{r=1}^{n}\sum_{s=1}^{n}\sin\frac{rs\pi}{n+1}\sin\frac{r\sigma\pi}{n+1}\,Y_s(t) \tag{7-20}$$

$$= \frac{n+1}{2}\,Y_\sigma(t)$$

Here again, use is made of the fact that

$$\left.\begin{aligned} \sum_{r=1}^{n} \sin\frac{rs\pi}{n+1}\sin\frac{r\sigma\pi}{n+1} &= 0 \qquad \text{when } s \neq \sigma \\ \sum_{r=1}^{n} \sin\frac{rs\pi}{n+1}\sin\frac{r\sigma\pi}{n+1} &= \frac{n+1}{2} \qquad \text{when } s = \sigma \end{aligned}\right\} \tag{7-21}$$

Problem 8. Illustrate equations (7-21) for the case $n = 3$.

Problem 9. Prove equations (7-21).

4. Forced Vibrations of a Loaded String.—If one of the masses on a string is subject to an external force, the motion of the system can still be described in terms of the response of the normal coordinates. Let particle p be acted on by a force $F_p \sin \omega t$. The differential equation containing d^2y_p/dt^2 will contain this force, but the others will remain as in equation (7-1). If now each equation is multiplied by $\sin rs\pi/(n+1)$ and they are added together, the result is

$$\frac{n+1}{2}\left(\frac{d^2Y_s}{dt^2} + \omega_s{}^2 Y_s\right) = F_p \sin\frac{ps\pi}{n+1}\sin\omega t \tag{7-22}$$

The solution of this equation was worked out in detail in Chap. IV. It can be concluded immediately that the response of a normal coordinate Y_s to a force applied to particle p is proportional to $\sin ps\pi/(n+1)$. In the case of three particles and a force applied to the center one, the normal coordinate with $s = 2$ will not respond at all.

Problem 10. Consider three particles on a string with a sinusoidal force applied to the central particle. Find the motion of each of the particles as a function of the frequency of the force, and show that there is a frequency at which the central particle does not move at all.

5. Approximation to a Continuous String.—If the length of the string is held fixed and the number of masses is increased, the loaded string approaches the string in which the mass is uniformly distributed. The various quantities pertaining to the loaded string also approach the corresponding quantities for the uniform string. Equation (7-15) for the displacement of a

particle can be expressed in terms of the distance of the particle from the end of the string, rather than in terms of the number r of the particle. This distance is $x = ra$, and the length of the string is $(n + 1)a$. Furthermore, the time dependence of the normal coordinates can be expressed in terms of an amplitude and a phase angle so that

$$y(x,t) = \sum_{s=1}^{n} A^{(s)} \sin \frac{s\pi x}{L} \cos (\omega_s t - \epsilon_s) \tag{7-23}$$

Although this expression has a definite magnitude for all values of x, it has significance for the loaded string only when x represents the position of a particle.

Problem 11. Show that ω_s is proportional to s when $s/(n + 1)$ is small.

Problem 12. Show that if the length of the string is held constant and n is increased, equation (7-1) approaches

$$\frac{\partial^2 y}{\partial t^2} = \frac{T}{\rho} \frac{\partial^2 y}{\partial x^2} \tag{7-24}$$

where ρ is the mass per unit length and y is the displacement, as a function of the time, of a point on the string whose coordinate is x.

6. Normal Vibrations of a Continuous String with Fixed Ends.—The above problems show the way in which both the differential equation and the solution for the loaded string approach those for the continuous string as the number of particles increases. It is possible also to treat the string as continuous from the beginning. By considering the force on an element of length as the sum of the tensions from the elements on either side, the equation of motion can be derived in the form of equation (7-24). The procedure in this derivation is really no different from that of obtaining the equation for the loaded string and passing to the limit. The use of elements and derivatives is essentially this process of passing to the limit. This one partial differential equation is equivalent to the infinity of ordinary differential equations that would be necessary if a different equation were written for each point

on the string; and in fact it is this infinite set, since for each value of x it is an ordinary differential equation in y as a function of t.

In solving the problem with discrete particles, a substitution was made to find a solution in which all the particles would vibrate with the same frequency but with different amplitudes. The sum of all such particular solutions, each multiplied by an arbitrary constant, then gave the general solution. The procedure for solving the partial differential equation is just the same. In this case, however, the different amplitudes must be expressed as functions of the coordinate x rather than of a discrete index. Hence the substitution will be made in the form

$$y = f(x)\phi(t) \tag{7-25}$$

When this is substituted in the partial differential equation, the result is the equation

$$\frac{\phi''}{\phi} = \frac{T}{\rho}\frac{f''}{f} \tag{7-26}$$

The left-hand side of this is a function of t only, while the right-hand side is a function of x only. These two sides can be equal to each other for all values of t and all values of x, only if each side is a constant. Let this constant be $-p$. Equation (7-26) is then equivalent to the two equations

$$\frac{d^2\phi}{dt^2} = -p\phi \qquad \text{and} \qquad \frac{d^2f}{dx^2} = -p\frac{\rho}{T}f \tag{7-26a}$$

The solutions of these equations are simply those for simple-harmonic motion. When they are multiplied together to give the expression for y, the result is

$$y = \left(C_0 \sin \sqrt{\frac{p\rho}{T}}\, x + D_0 \cos \sqrt{\frac{p\rho}{T}}\, x\right) (A_0 \sin \sqrt{p}t + B_0 \cos \sqrt{p}t) \tag{7-27}$$

Equation (7-27) gives a value of y that will satisfy the partial differential equation for any values of the five constants A_0, B_0, C_0, D_0, and p. Of the first four, only three are independent

as can be seen by combining the sine and cosine into a single trigonometric function with a phase constant. However, in addition to the differential equation, it is necessary to satisfy the boundary conditions. The problem provides that the ends of the string are fixed so that their displacement is always zero. Hence, $y = 0$ for $x = 0$ and $x = L$, for all values of the time. This requires then, either that ϕ shall be zero at all times, which is a trivial case, or else that $f(x)$ shall be zero at the end points. When $x = 0$, $\sin \sqrt{p\rho/T}x$ is zero for all values of p, but $\cos \sqrt{p\rho/T}x$ is not zero for any value of p. Hence the boundary conditions can be satisfied only by making $D_0 = 0$. To satisfy the condition at the other end, it is necessary that $\sin \sqrt{p\rho/T}L = 0$. This is true only for certain values of p, which are given by the equation

$$\frac{p\rho}{T} = \frac{n^2\pi^2}{L^2} \tag{7-28}$$

where n is an integer. The second factor in (7-27) shows that these values of p determine the frequencies with which the system can vibrate. Hence this continous system can vibrate with only a discrete set of normal frequencies just as systems of discrete particles can vibrate with only certain normal frequencies. The final result can be written

$$y_n = \sin \frac{n\pi x}{L}\left(A_n \sin \sqrt{\frac{T}{\rho}}\frac{n\pi}{L}\, t + B_n \cos \sqrt{\frac{T}{\rho}}\frac{n\pi}{L}\, t\right) \tag{7-29}$$

where A_n and B_n are the arbitrary constants that determine the amplitude and the phase of the vibration.

Equation (7-29) gives a particular solution of the problem. In addition to being a particular solution this is also a solution that represents a normal vibration, since all parts of the string vibrate in phase. The time-dependent part of (7-29), or the function ϕ_n, may be considered as a normal coordinate of the system. Each normal coordinate is associated with one of the permitted values of p and may be designated by the integer n that fixes p. Since there is no limit to the number of possible values of p, there is no limit to the number of possible normal

coordinates. This is as it should be, since there is an infinity of particles in the string and an infinity of coordinates is necessary in order to locate them all. These normal coordinates have the properties previously tabulated for such quantities. The differential equations governing their variation are those for simple-harmonic motion, such as the first of (7-26*a*). It is also possible to set one into motion independently of the others as in the solution (7-29). The manner in which the energy of a string can be expressed in terms of the normal coordinates will be treated later.

7. General Solution and Evaluation of Constants.—Equation (7-29) is a particular integral of the partial differential equation that describes the motion of the string. To get the general solution it is necessary to add all the particular solutions, each multiplied by suitable arbitrary constants. These arbitrary constants are the A_n and the B_n of equation (7-29). Hence the general solution is

$$y = \sum_{n=1}^{\infty} \sin \frac{n\pi x}{L} \left(A_n \sin \sqrt{\frac{\overline{T}}{\rho}} \frac{n\pi}{L} t + B_n \cos \sqrt{\frac{\overline{T}}{\rho}} \frac{n\pi}{L} t \right) \quad (7\text{-}30)$$

This solution contains an infinity of arbitrary constants, which must be determined from the initial conditions. As initial conditions there can be given the position and the velocity of each point. These can best be given as functions of x. Suppose that, at $t = 0$, $y_0 = g(x)$ and $\dot{y}_0 = h(x)$. It then follows from (7-30) that

$$\left.\begin{aligned} g(x) &= \sum_{n=1}^{\infty} B_n \sin \frac{n\pi x}{L} \\ \text{and} \qquad h(x) &= \sqrt{\frac{\overline{T}}{\rho}} \frac{\pi}{L} \sum_{n=1}^{\infty} n A_n \sin \frac{n\pi x}{L} \end{aligned}\right\} \quad (7\text{-}31)$$

If it is possible to find such values of A_n and B_n that these equations are satisfied, the arbitrary constants will have been determined.

The problem of finding the coefficients is thus reduced to the problem of expressing the arbitrary functions $g(x)$ and $h(x)$ as sums of sines and of finding the coefficients in these sums. These functions g and h are also subject to the condition that they are zero at $x = 0$ and $x = L$. The theory of Fourier series gives the method for finding the coefficients. To find the coefficient B_r, multiply both sides of the first of equations (7-31) by $\pi/L \sin r\pi x/L\, dx$, and integrate between 0 and L.

$$\frac{\pi}{L}\int_0^L g(x) \sin\frac{r\pi x}{L}\,dx = \frac{\pi}{L}\sum_{n=1}^{\infty} B_n \int_0^L \sin\frac{n\pi x}{L}\sin\frac{r\pi x}{L}\,dx = \frac{\pi}{2}B_r$$

This gives

$$B_r = \frac{2}{L}\int_0^L g(x) \sin\frac{r\pi x}{L}\,dx \tag{7-32}$$

The values of A_n can be determined in a similar way.

Problem 13. Explain and justify the steps leading to equation (7-32).

Problem 14. Evaluate the constants A_n in equation (7-30) in terms of $g(x)$ and $h(x)$.

Problem 15. A stretched string is displaced to the position given by $y = D \sin 3\pi x/L$ and let go with zero initial velocity. Find the subsequent motion. Is this a normal vibration?

Problem 16. A stretched string is pulled aside at the center to a distance d and then let go. Find the subsequent motion.

Problem 17. A string is struck so as to give an initial velocity v to the central portion for a distance s on each side of the center. Find the subsequent motion.

8. The Energy of a Vibrating String.—The kinetic energy of a vibrating string is simply the sum of the energies of all the elements of length. Hence

$$2T = \int_0^L \rho\left(\frac{\partial y}{\partial t}\right)^2 dx \tag{7-33}$$

The potential energy can be obtained as the limit of the expression for the potential energy of the loaded string, as the number of particles increases. To find the energy of the loaded

string, it is convenient to consider the string first in the equilibrium position. Then the first particle is moved to its desired position, and all the other particles are moved at the same time to a displacement equal to that of the first particle. The work done in making these displacements can easily be determined. The only force on the first particle is due to the first segment of string, while the only force on the last particle is due to the last segment of string. The other particles experience no force. The second to the last particles are then moved to the desired position of the second particle. In this process, only the second and the last segments of string exert any force, and the work done can be easily calculated. By a continuation of this process the particles can all be put into their desired positions, and the amount of work necessary can be calculated. The result gives for the potential energy

$$V = \frac{T}{2a} \sum_{r=1}^{n+1} (y_r - y_{r-1})^2 \tag{7-34}$$

In this expression T is the tension, y_0 represents the displacement of one fixed end of the string, which is, of course, zero, while y_{n+1} represents the displacement of the other end of the string, which is also always zero. The inclusion of these two zero displacements makes the notation a little simpler.

In the limit in which the number of particles becomes infinite and the string becomes uniformly loaded, equation (7-34) becomes

$$V = \frac{T}{2} \int_0^L \left(\frac{\partial y}{\partial x}\right)^2 dx \tag{7-34a}$$

Problem 18. Express the potential energy of a loaded string as given in equation (7-34) in terms of normal coordinates Y_s.

Problem 19. Express the potential energy of a continuous stretched string in terms of the normal coordinates ϕ_n.

9. Forced Vibration of a Continuous String.—The problem of forced vibrations can be treated by analogy with the case of a loaded string and in terms of normal coordinates. Let a sinusoidal force be applied to a string with an amplitude that depends

on position. It may then be written $F(x) \sin \omega t$. The function $F(x)$ specifies the way in which the force is distributed over the string. For a force applied at a single point, $F(x)$ is zero except at the point of application. If a more complicated dependence on the time is involved, it can be built up of a series of sine functions.

The differential equation of motion of the string is then

$$\rho \frac{\partial^2 y}{\partial t^2} - T \frac{\partial^2 y}{\partial x^2} = F(x) \sin \omega t \tag{7-35}$$

If the solution is to be expressed in terms of the normal coordinates,

$$y = \sum_{n=1}^{\infty} \sin \frac{n\pi x}{L} \phi_n \tag{7-35a}$$

Substituted in the differential equation, this leads to

$$\sum_{n=1}^{\infty} \left(\rho \ddot{\phi}_n + \frac{n^2\pi^2}{L^2} T \phi_n \right) \sin \frac{n\pi x}{L} = F(x) \sin \omega t \tag{7-35b}$$

If now both sides of the equation are multiplied by $\sin s\pi x/L$ and integrated from $x = 0$ to $x = L$, only one term of the sum gives an integral different from zero and it follows that

$$\rho \ddot{\phi}_s + \frac{s^2\pi^2}{L^2} T \phi_s = \frac{2}{L} \int_0^L F(x) \sin \frac{s\pi x}{L} \, dx \sin \omega t \tag{7-36}$$

This shows that the normal coordinate ϕ_s will be forced to oscillate with the frequency ω. Damping has not been considered in this problem, but it will have the usual effect of causing the transient motion to die out and the steady state to be only an oscillation with the frequency of the force. The steady-state oscillation of each normal coordinate will have an amplitude proportional to the integral on the right-hand side of equation (7-36), which is just the coefficient in the Fourier-series expansion of the function $F(x)$.

Problem 20. Find the steady-state motion of a stretched string acted on by a force distributed as $\sin \pi x/L$ and varying sinusoidally

with a frequency much lower than the fundamental natural frequency ω_1.

Problem 21. Show that a force applied at a point P of a string will not stimulate normal vibrations having a node at P.

10. Expansion of Functions as Series of Orthogonal Functions.—According to equation (7-30) the shape of a string at any time is given by a series of sines of multiples of x in which the coefficients are functions of the time. At the particular time $t = 0$ the coefficients are simply the B's. At other times the coefficients are combinations of the A's and the B's. The significance of this statement is that any function $y(x)$, which represents a possible displacement and position of the string, can be represented as a series of sine functions of x. This is merely a special case of the expansion of functions in series of orthogonal functions. The evaluation of the constants in such a series depends upon the property of the orthogonality of certain groups of functions.

Two functions $R_1(x)$ and $R_2(x)$ are said to be orthogonal to each other in a certain range of the independent variable $a \leq x \leq b$ if $\int_a^b R_1 R_2 \, dx = 0$. If in a certain set of functions $R_1, R_2, \ldots, R_n, \ldots$, each of the functions is orthogonal to all the rest, the functions are said to form an *orthogonal set.* If there exists no function, except zero, that is orthogonal to all the members of the set, the set is said to be a *complete orthogonal set.* With certain restrictions, which are usually of little importance in physics, any arbitrary function can be expressed as a series of the functions of a complete orthogonal set, with suitable coefficients.

A function can be normalized by multiplying it by a constant. When it is normalized $\int_a^b R_i^2 \, dx = 1$. Hence a set of normalized orthogonal functions has the property that

$$\int_a^b R_i R_j \, dx = \delta_{ij} \tag{7-37}$$

The quantity δ_{ij} is a function of the subscripts. It is equal to 1 when $i = j$ and is equal to 0 otherwise.

The orthogonality is essential to the determination of the coefficients in a series, while the normalization is convenient. Let

$$F(x) = A_1R_1 + A_2R_2 + \cdots A_nR_n + \cdots \quad (7\text{-}38)$$

where the R's are the members of a complete set of normalized orthogonal functions. To determine the value of the coefficient A_n, multiply both sides of this equation by R_n, and integrate between the limits a and b. Then, because of the normalization and the orthogonality,

$$A_n = \int_a^b R_n(x)F(x)dx \quad (7\text{-}39)$$

Thus the coefficients can always be determined by a single integration. The investigation of the convergence of these series is too elaborate to be undertaken here.

Between the limits 0 and π the functions $\sin nx$, with n integral, form a complete orthogonal set. The functions $\cos nx$ form a similar set. Between the limits $-\pi$ and π, the combination of these two sets gives a complete orthogonal set. A series of such functions, valid between $-\pi$ and π is called a *Fourier series*, after its inventor.

There are many sets of orthogonal functions. One is formed of the functions $H_ne^{-x^2/2}$, where the H_n's are the Hermite polynomials. These polynomials are

$$H_n = (2x)^n - \frac{n(n-1)}{1!}(2x)^{n-2} + \frac{n(n-1)(n-2)(n-3)}{2!}(2x)^{n-4} \cdots \quad (7\text{-}40)$$

As given in this form the functions are not normalized. They can be normalized by dividing by the square root of the integral of the square of the function from negative to positive infinity. These functions are orthogonal for the whole range of the variable x from $-\infty$ to $+\infty$.

Problem 22. Expand the function $y = x$, in a series of sines between $x = 0$ and $x = \pi$.

Problem 23. If the function $y(x)$ is such that $y = 0$ for values of x between $-\pi$ and 0 and $y = x$ for values of x between 0 and π, find the Fourier series for $y(x)$.

Problem 24. Normalize the orthogonal functions that involve the Hermite polynomials for $n = 0$, 1, and 2. Find the first two coefficients in the expansion of $y = e^{-x^2}$ in terms of these functions.

11. Vibration of a Nonuniform String.—The problem of determining the vibration of a nonuniform string may be taken as an example of a large class of vibration problems that are more complicated than those thus far treated but about which something can be learned by comparison with the simpler problems. The partial differential equation of the motion is

$$\rho(x) \frac{\partial^2 y}{\partial t^2} = T \frac{\partial^2 y}{\partial x^2} \tag{7-41}$$

This differs from equation (7-24) in that the density ρ is not a constant but is a function of x. To find the normal vibrations one proceeds in the usual way with the substitution $y = f(x)\phi(t)$. Substitution in the differential equations and the separation of the variables give

$$\frac{d^2\phi}{dt^2} = -p\phi \tag{7-41a}$$

$$\frac{d^2 f}{dx^2} = -\frac{p}{T}\rho(x) f \tag{7-41b}$$

The period of vibration of ϕ is given by the constant p, and p must have such a value that there exists a function $f(x)$ which satisfies the differential equation (7-41b) as well as the boundary conditions. Equation (7-41b) is an example of the Sturm-Liouville type of equation. It has no simple solution for an arbitrary density $\rho(x)$, but a number of its properties can be determined with relative ease.

Since the density is always positive, the differential equation shows that the second derivative of f will always have the opposite sign to f itself. This means that the curve will always be concave toward the axis so that the function will cross the axis at a number of points. If $p = 0$, the only value of f that can satisfy the boundary conditions is $y = 0$. This is, however, a

trivial solution. If then the value of p is allowed to increase, the roots of the function come closer and closer together until, if the function starts from the origin, the first positive root coincides with the other end of the string. This is then the smallest value of p that satisfies the conditions of the problem. It gives the frequency of the fundamental vibration. As p is again allowed to increase, there comes a value for which the second root of f will coincide with the end of the string. The process can be continued indefinitely to show that there is an infinity of discrete values of p which satisfy the conditions of the problem. In principle they could be determined by graphical integration and interpolation.

It is also possible to show that the various functions $f(x)$, when multiplied by $\rho^{\frac{1}{2}}(x)$, form an orthogonal set. Let f_i be the function associated with the value p_i of p, and let f_j be the function associated with the value p_j. The differential equation for f_i is multiplied by f_j, the differential equation for f_j is multiplied by f_i, and the difference between the two equations is integrated between $x = 0$ and $x = L$. The integral of the two terms containing second derivatives is equal to zero so that

$$(p_i - p_j) \int_0^L \rho(x) f_i(x) f_j(x) dx = 0 \tag{7-42}$$

This shows that the functions are orthogonal unless $p_i = p_j$. For this case the integral can be made equal to 1 by suitable normalization. This is a general method of proving the orthogonality of functions from their differential equations.

Problem 25. Carry through in detail the proof of equation (7-42).

Problem 26. Use the method just indicated to show that the sines and cosines of integral multiples of the independent variable are orthogonal functions between $-\pi$ and π.

Problem 27. The differential equation for the Hermite polynomials is

$$\frac{d^2H_n}{dx^2} - 2x\frac{dH_n}{dx} + 2nH_n = 0 \tag{7-43}$$

From this show that the functions $H_n e^{-x^2/2}$ form an orthogonal set.

Since the functions $f_j(x)$ in equation (7-41b) must be zero at both ends of the string, they are adapted to an expansion in a series of sines. Let

$$f_j(x) = \sum_{r=1}^{\infty} A_r{}^j \sin \frac{r\pi x}{L} \tag{7-44}$$

The coefficients in this expansion must then be determined from the differential equation. When the series is substituted, the result is

$$\sum_{r=1}^{\infty} A_r{}^j \left(\frac{r^2\pi^2}{L^2}\right) \sin \frac{r\pi x}{L} - \frac{p}{T} \sum_{r=1}^{\infty} A_r{}^j \rho(x) \sin \frac{r\pi x}{L} = 0 \tag{7-44a}$$

If this equation is multiplied by $\sin s\pi x/L$ and then integrated from $x = 0$ to $x = L$, the result is

$$A_s{}^j \left(\frac{s^2\pi^2}{2L}\right) - \frac{p}{T} \sum_{r=1}^{\infty} A_r{}^j \int_0^L \rho(x) \sin \frac{r\pi x}{L} \sin \frac{s\pi x}{L}\, dx = 0 \tag{7-44b}$$

An equation of the form of (7-44b) can be obtained for every value of s, so that one has a set of simultaneous equations for the determination of the constants $A_r{}^j$. These equations will be compatible only for those values of p that make the determinant of the coefficients equal to zero. Each value of p, say p_j, has associated with it a set of values of the coefficients designated by the superscript j. Since there is an infinity of equations and each equation contains an infinity of terms, the determinant has an infinite number of rows and columns. Although a general treatment is not possible, it is possible to find approximate solutions for certain special cases.

In some cases the density can be expressed as a constant density plus a variable part. Then, if $\rho(x) = \rho_0 + \sigma(x)$, equation (7-44$b$) becomes

$$A_s{}^j \left(\frac{s^2\pi^2}{2L} - \frac{pL\rho_0}{2T}\right) - \frac{p}{T} \sum_{r=1}^{\infty} A_r{}^j \int_0^L \sigma(x) \sin \frac{r\pi x}{L} \sin \frac{s\pi x}{L}\, dx = 0 \tag{7-44c}$$

If $\sigma(x)$ is small compared with ρ_0, an approximation to the frequencies can be obtained by neglecting all except the diagonal terms of the determinant. This is equivalent to neglecting quantities of the order of $(\sigma/\rho)^2$.

Problem 28. A small mass m is attached to the middle point of a vibrating string. Find the first approximation to the frequencies of vibration.

Problem 29. Show that a small mass attached to any point on a vibrating string will not affect the frequencies or the forms of those normal modes of vibration which have a node at the point of attachment.

12. The Variation Problem for Normal Vibrations.—Some useful results can be obtained from the treatment of equation (7-41*b*) as the differential equation for a variation problem. If it is required to find the form of $f(x)$ that gives the integral

$$I = \int_0^L \left(\frac{df}{dx}\right)^2 dx \tag{7-45}$$

a stationary value, while

$$S = \frac{1}{T}\int_0^L \rho(x) f^2\, dx = 1 \tag{7-45a}$$

the Lagrangian equation becomes just equation (7-41*b*). Hence, if some other method of solving the variation problem can be found, the result is a solution of the differential equation.

Furthermore, by multiplying equation (7-41*b*) by $f(x)$ and integrating between the limits $x = 0$ and $x = L$, it can be shown that the integral I gives the value of p.

Although it is not easy to get an exact solution of the variation problem, it is possible to get an upper limit for the smallest value of p by this method. Since the lowest value of p corresponds to that function $f(x)$ which makes the integral in (7-45) a minimum, it is evident that the substitution of any other function in the integral will give a value of p greater than the true minimum value. Hence, it is possible by the trial of several functions to find the one that gives the lowest and, hence, best value of p. One method is to take a function that

depends upon one or more parameters, and then by differentiation with respect to the parameters to find those values which give the lowest frequency. By a judicious selection of the functions taken for trial, a very good approximation can be obtained to the lowest or fundamental frequency.

Problem 30. Show that the Lagrangian equation for the variation problem of equations (7-45) and (7-45*a*) is the differential equation (7-41*b*).

Problem 31. Show that the integral I of equation (7-45) is equal to the corresponding value of p.

Problem 32. Consider a stretched string with a mass m attached to its center point. Assume that the function $f(x)$ for the fundamental mode of vibration has the form

$$f(x) = 1 - \left(\frac{2x}{L}\right)^n$$

on the positive side of the origin, which is taken at the center, and is symmetrical about the center. Find the value of n that gives the best approximation to the frequency. Compare this result with the exact result in the case where $m = 0$.

13. Traveling Waves in a String.—Thus far the problem of a vibrating string has been treated from the point of view of normal coordinates, which represent a vibration of the whole string. It is also possible to treat the propagation of disturbances along the string. The general solution of the partial differential equation (7-24) can be written

$$y = F_1(x - vt) + F_2(x + vt) \tag{7-46}$$

This solution is similar to that in equation (7-27) since it satisfies no particular boundary conditions. The constant v is equal to $(T/\rho)^{\frac{1}{2}}$, while F_1 and F_2 are arbitrary functions of their respective arguments. The generality of the solution is indicated by the appearance of arbitrary functions instead of merely arbitrary constants such as occur in the solutions of ordinary differential equations. The solution represents a disturbance of the form of F_1 moving toward the positive values of x with the velocity v and another disturbance of the form of F_2 moving with

the same velocity in the opposite direction. If the ends of the string are so far away that the disturbances cannot reach them during the time in which the behavior of the string is observed, the ends may be neglected and the string treated as infinite. The description in terms of traveling waves is more suited to this case than that of normal vibrations.

The traveling-wave point of view can also be used when boundary conditions must be satisfied. If it is required that the string be fastened at $x = 0$, the solution (7-46) must be subjected to the restriction that $F_1(-vt) = -F_2(vt)$ for all values of t. If, in addition, it is required that the other end of the string at $x = L$ is to be fixed, the functions F_1 and F_2 must be periodic with the period L. When subject to these conditions, the general solution is such that it can be interpreted as representing a disturbance reflected back and forth between the ends of the string.

Although the functions F_1 and F_2 are perfectly arbitrary, they can be expressed, over any finite range of the independent variable, as Fourier series. On this account it is sufficient for many purposes to study the behavior of sinusoidal waves of the form

$$y = \sin \frac{2\pi}{\lambda}(x - vt) \qquad \text{and} \qquad y = \sin \frac{2\pi}{\lambda}(x + vt) \quad (7\text{-}47)$$

The constant λ is called the *wave length*, since at a given time, if x changes by λ, the argument of the sine changes by 2π. It is clear in a similar way that v/λ is the frequency with which any point of the string vibrates.

Problem 33. Show that in a string of length L the solution

$$y = A \sin \frac{n\pi}{L}(x - vt) + A \sin \frac{n\pi}{L}(x + vt)$$

where n is an integer, satisfies the boundary conditions. Express this solution in the form of equation (7-30), and determine whether it represents a normal vibration.

Problem 34. Consider a stretched string of infinite length whose density has one constant value for negative values of x and another

constant value for positive values of x. Assume that a train of waves of the form (7-47) comes from the negative x axis and is partly reflected and partly transmitted at the discontinuity. By applying the conditions arising from the continuity of the string, match the solutions on the two sides of the origin, and determine the relative amplitudes of the transmitted and the reflected waves.

References

RAYLEIGH: "Theory of Sound," The Macmillan Company, New York.

SLATER and FRANK: "Introduction to Theoretical Physics," Chaps. 11 to 15, McGraw-Hill Book Company, Inc., New York.

WEBSTER, A. G.: "Partial Differential Equation of Mathematical Physics," B. G. Teubner, Leipzig.

CHAPTER VIII

VECTOR ANALYSIS

In the study of a physical problem it is usually necessary to establish a system of coordinates in which the various material bodies can be located. The positions of the bodies, expressed in terms of these coordinates, are then the variables in the equations that state the laws of physics. Although these equations contain explicit reference to the coordinates used, it is difficult to believe that the relationships expressed should depend upon the coordinates in any essential way. The laws of physics should be relationships between physical things, and these relationships should be true regardless of the language used to state them. In fact, it has long been taken as an axiom that the laws of physics must be expressible in a form that is the same for all systems of coordinates.

The treatment of Hamilton's principle, in Chap. VI, has shown one method of expressing the Newtonian laws of motion in a form independent of any particular system of coordinates. The use of vectors is another way in which this object may be attained. A vector is independent of any particular coordinate system and can be used for the representation of physical quantities.

1. The Definition of a Vector.—Many physical quantities are of such a nature that they cannot be completely specified by single numbers but require in addition the specification of directions. For example, although a mass is completely given by a single number, a force is not completely defined until both its magnitude and direction are known. Quantities of this kind, such as force, velocity, or momentum, are called *vector quantities*. Quantities such as mass are called *scalar quantities*.

A vector can be defined as a line whose length and direction represent the magnitude and direction of a vector quantity. To

represent the direction unambiguously it is necessary to distinguish between the origin and the end of the vector. A vector will be regarded as unchanged when it is moved parallel to itself. If the point of application of a force is of importance, this is regarded as an additional specification, and not as a property of the vector.

A vector can be multiplied by a scalar to give a vector whose direction is that of the original vector but whose magnitude is the original magnitude multiplied by the scalar. Multiplication by -1 interchanges the origin and the end. Multiplication by a negative scalar is equivalent to multiplication by -1 and then by the absolute value of the scalar.

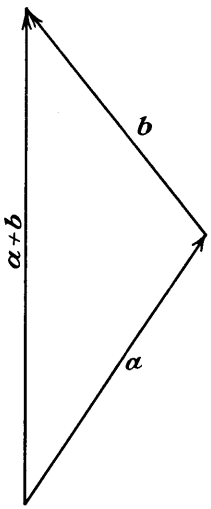


FIG. 8-1.—Addition of the vectors **a** and **b** to form the sum **a** + **b**.

Two vectors are equal when they have the same magnitude and the same direction. A vector equation always implies this kind of equality between its two sides and so contains more information than does a scalar equation. If a vector is equal to zero, its length is zero and its direction is indeterminate.

Vectors will be designated by letters in boldface type, and the magnitude of the vector will be represented by the same letter in italics. Thus $x\mathbf{a}$ designates a vector whose direction is that of **a** and whose magnitude is xa.

2. Addition of Vectors.—The addition of vectors is so defined as to satisfy as many as possible of the laws of addition of scalars and also to be useful in representing physical relationships. To add two vectors the origin of the second vector is placed at the end of the first. The sum is then the vector whose origin coincides with the origin of the first vector and whose end coincides with the end of the second. Figure 8-1 illustrates this relationship. These three vectors clearly lie in the same plane.

A number of geometrical theorems can be established by means of vector algebra, using only addition and subtraction of vectors and multiplication by scalars. Consider, for example,

the intersection of the diagonals of a parallelogram. In Fig. 8-2 the sides of the parallelogram are designated by **a** and **b** so that the diagonals are **a** + **b** and **a** − **b**. The point of intersection is a fraction, say x, of the distance along **a** + **b** and another fraction, say y, of the distance along **a** − **b**. Hence

$$x(\mathbf{a} + \mathbf{b}) = \mathbf{b} + y(\mathbf{a} - \mathbf{b})$$

or

$$(x - y)\mathbf{a} = (1 - x - y)\mathbf{b} \tag{8-1}$$

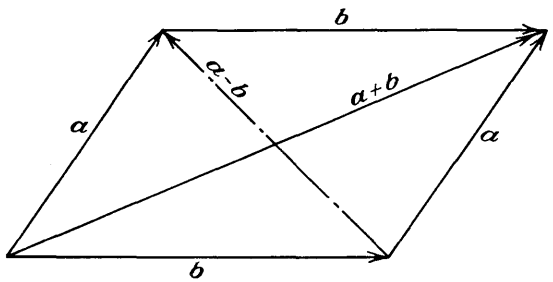


FIG. 8-2.—The sides and diagonals of a parallelogram expressed in terms of the vectors **a** and **b**.

Since the vectors **a** and **b** are not parallel, this equation can be satisfied only when both sides are zero. This requires

$$x = y = \tfrac{1}{2}$$

and thus the diagonals bisect each other.

Problem 1. If $x\mathbf{a} + y\mathbf{b} = 0$ and $x + y = 0$, show that **a** and **b** have the same length and the same direction.

Problem 2. Show that the middle points of the lines joining the middle points of the opposite sides of a quadrilateral coincide, whether the four sides are in the same plane or not.

Problem 3. Show that the line drawn from one vertex of a parallelogram so as to trisect the diagonal bisects the opposite sides.

3. Orthogonal Components of Vectors.—In a system of Cartesian coordinates, any vector can be represented as the sum of three vectors lying along the coordinate axes. Let **i**, **j**, and **k** be vectors of unit length lying along the x, y, and z axes, respectively. Then any vector **a** can be written

$$\mathbf{a} = a_x\mathbf{i} + a_y\mathbf{j} + a_z\mathbf{k} \tag{8-2}$$

a_x, a_y, and a_z are scalar quantities called the *scalar components* of **a**, and the vectors $a_x\mathbf{i}$, $a_y\mathbf{j}$, and $a_z\mathbf{k}$ may be called the *vector components* of **a**. These components are unique. No other triad of numbers will serve as components of this vector along these three directions.

A vector can be expressed as the sum of components along any three noncoplanar axes and hence in terms of any three noncoplanar unit vectors. The unit vectors **i**, **j**, and **k**, however, are especially useful because of their orthogonality. For practical computation with vectors it is nearly always necessary to make use of these rectangular components.

Problem 4. Write the vector equation for the center of mass of a number of particles. The vector representing a point is the vector drawn from the origin of coordinates to that point. Show also that the center of mass, defined in this way, is independent of the point chosen as origin.

Problem 5. A group of forces acting at the point O are represented by the vectors OA, OB, OC, . . . , ON. Show that, if the forces are in equilibrium, O is the centroid of the points A, B, C, . . . , N. A centroid is defined so that it is the center of mass, if unit masses are placed at all of the points in question.

Problem 6. A person who is moving eastward at 3 miles per hour finds that the wind appears to blow directly from the north. On doubling his speed, it appears to come from the northeast. Find the vector wind velocity.

Problem 7. If three vectors **a**, **b**, and **c** start from the same point and end on the same straight line, find a relationship between them.

4. Multiplication of Vectors. *a. The Scalar Product.*—It has been found useful to define two kinds of vector multiplication. The first kind gives the *scalar product.* The scalar product of two vectors is equal to the product of their lengths and the cosine of the angle between them.

$$\mathbf{a} \cdot \mathbf{b} = ab \cos (\mathbf{ab}) \tag{8-3}$$

In Gibbs' notation it is denoted by a dot between the letters denoting the vectors and hence is frequently called the *dot product.* Another notation, frequently used in Europe, encloses the two letters in parentheses.

The scalar product is of much use in physics. For example, the work done by a force during a displacement is equal to the scalar product of the force and the displacement.

It must be noted that the scalar product of two vectors is a scalar, not another vector. From the definition it is obvious that this product is commutative; the order of the factors is immaterial. The distributive law is also obeyed.

Problem 8. Find the scalar products of the unit vectors **i**, **j**, and **k** with each other.

Problem 9. Find the scalar product of an aribtrary vector **a** and the unit vectors along the coordinates axes.

Problem 10. Show that the scalar product of two vectors, in terms of their components, is

$$\mathbf{a} \cdot \mathbf{b} = a_x b_x + a_y b_y + a_z b_z \tag{8-4}$$

Problem 11. Show that the two vectors $\mathbf{a} = l\mathbf{i} + m\mathbf{j} + n\mathbf{k}$ and $\mathbf{b} = \lambda\mathbf{i} + \mu\mathbf{j} - [(l\lambda/n) + (m\mu/n)]\mathbf{k}$ are perpendicular to each other.

Problem 12. Show that the sum of the squares of the diagonals of a parallelogram is equal to the sum of the squares of the sides.

b. The Vector Product.—The other kind of product of two vectors is called the *vector product.* In Gibbs' notation it is indicated by a small cross between the letters representing the vectors, and therefore it is often called the *cross product.* The definition of the vector product is given by the equation

$$\mathbf{a} \times \mathbf{b} = ab \sin (\mathbf{ab})\mathbf{e} \tag{8-5}$$

where **e** is a unit vector perpendicular to both **a** and **b.** The sense of **e** is such that a right-handed screw would advance in the positive direction of **e** if it were turned from **a** to **b.** From this definition, and the specification of **e,** it is clear that

$$\mathbf{a} \times \mathbf{b} = -\mathbf{b} \times \mathbf{a} \tag{8-6}$$

Thus the commutative law does not hold, and the order of the factors is important in this product. The distributive law, however, does hold for this kind of multiplication.

The result of vector multiplication is a vector and therefore has a direction as well as a magnitude. This distinction between scalar and vector products is of importance in forming the prod-

uct of three or more factors. It is impossible to form a scalar product of three factors, since the product of any two of them is already a scalar. The only possible interpretation of such a product would be to multiply the third vector by the scalar product of the other two. The result would then depend upon the two that were selected for forming the first product. On the other hand, it is possible to form a vector product of three vectors, although it is necessary to specify the association of the vectors.

Problem 13. Evaluate the products $\mathbf{i} \times \mathbf{i}$, $\mathbf{i} \times \mathbf{j}$, $\mathbf{i} \times \mathbf{k}$.

Problem 14. Show that the vector product of two vectors can be written in terms of their components, formally, as the determinant

$$\mathbf{a} \times \mathbf{b} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_x & a_y & a_z \\ b_x & b_y & b_z \end{vmatrix} \tag{8-7}$$

Problem 15. Find the area of the triangle bounded by the vectors **a**, **b**, and their difference.

Problem 16. Show that the vector product obeys the distributive law of ordinary multiplication.

Problem 17. Show that the volume of the parallelepiped whose edges are the vectors **a**, **b**, and **c** is given by the triple scalar product $\mathbf{a} \cdot \mathbf{b} \times \mathbf{c}$.

Problem 18. Show that

$$\mathbf{a} \cdot \mathbf{b} \times \mathbf{c} = \begin{vmatrix} a_x & a_y & a_z \\ b_x & b_y & b_z \\ c_x & c_y & c_z \end{vmatrix} \tag{8-8}$$

Problem 19. Show that

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = \mathbf{b}(\mathbf{a} \cdot \mathbf{c}) - \mathbf{c}(\mathbf{a} \cdot \mathbf{b}) \tag{8-9}$$

Problem 20. Show that, if a line passes through the centroid of a number of points, the sum of the perpendiculars from the points to this line is zero. These perpendiculars are to be treated as vectors whose origins lie on the line through the centroid and whose ends are at the various points.

Problem 21. Show that the work done by a force during the displacement of its point of application is equal to the sum of the quantities of work done by the components of the force.

Problem 22. A fluid is flowing through a plane surface with a uniform velocity **q**. If **n** is the unit normal to the plane, show that the

volume of the fluid passing through unit area of the plane per unit time is $\mathbf{q} \cdot \mathbf{n}$.

5. Differentiation of a Vector with Respect to a Scalar.—If a vector $\mathbf{a}$ is a function of the scalar variable t, then to every value of t will correspond a certain length and direction of $\mathbf{a}$. If the value of t is changed slightly, the vector $\mathbf{a}$ will be changed slightly and in general both the length and the direction will be changed. If the difference between the two values of t is Δt, and the difference between the two corresponding values of $\mathbf{a}$ is $\Delta \mathbf{a}$, then the derivative of a with respect to t is

$$\frac{d\mathbf{a}}{dt} = \lim_{t \to 0} \frac{\Delta \mathbf{a}}{\Delta t} \tag{8-10}$$

if this limit exists. This derivative is evidently a vector and therefore has a direction as well as a magnitude. If the vectors are referred to a fixed set of rectangular axes, the derivative can be expressed in terms of its components.

$$\frac{d\mathbf{a}}{dt} = \mathbf{i} \frac{da_x}{dt} + \mathbf{j} \frac{da_y}{dt} + \mathbf{k} \frac{da_z}{dt} \tag{8-11}$$

It is important to note that this form of the derivative is valid only when the coordinate axes are fixed and do not change with the independent variable. In some problems it is convenient to use rotating axes, and in these cases the time derivative of a vector, when expressed in terms of its components, must contain terms representing the rates of change of the axes.

The extension of these statements to higher derivatives is relatively obvious. The differentiation of sums and products will be the same as for ordinary scalars, except that attention must be paid to the order of the factors in the case of vector products.

If $\mathbf{r}$ represents a vector from a fixed origin to a point moving in a plane, the rate of change of $\mathbf{r}$ is often conveniently expressed in terms of unit vectors along and perpendicular to its length. Let $\mathbf{r}_1$ be the unit vector in the direction of $\mathbf{r}$, and let $\boldsymbol{\theta}_1$ be a unit vector perpendicular to $\mathbf{r}$. Then $\mathbf{r} = r\mathbf{r}_1$, and

$$\frac{d\mathbf{r}}{dt} = \frac{dr}{dt} \mathbf{r}_1 + r \frac{d\mathbf{r}_1}{dt} \tag{8-12}$$

If $\mathbf{r}$ is confined to a plane and its direction is specified by the angle θ,

$$\frac{d\mathbf{r}_1}{dt} = \frac{d\theta}{dt}\,\boldsymbol{\theta}_1$$

whence

$$\frac{d\mathbf{r}}{dt} = \frac{dr}{dt}\,\mathbf{r}_1 + r\,\frac{d\theta}{dt}\,\boldsymbol{\theta}_1 \tag{8-13}$$

Problem 23. If $\mathbf{r}$ is any vector that is a function of the time t, and if $\mathbf{r} \cdot d\mathbf{r}/dt = 0$ at all times, show that the length of the vector is constant. Also show that, if $\mathbf{r} \times d\mathbf{r}/dt = 0$ at all times, the vector has a constant direction.

Problem 24. If $\mathbf{r}_1$ is a unit vector in the direction of $\mathbf{r}$, show that

$$\mathbf{r}_1 \times d\mathbf{r}_1 = \mathbf{r} \times d\mathbf{r}/r^2.$$

Problem 25. Show that

$$\frac{d^2\mathbf{r}}{dt^2} = \left[\frac{d^2r}{dt^2} - r\left(\frac{d\theta}{dt}\right)^2\right]\mathbf{r}_1 + \left(r\,\frac{d^2\theta}{dt^2} + 2\,\frac{dr}{dt}\,\frac{d\theta}{dt}\right)\boldsymbol{\theta}_1 \tag{8-14}$$

Problem 26. The equation $\mathbf{r} = \mathbf{a}\cos t + \mathbf{b}\sin t$, where $\mathbf{a}$ and $\mathbf{b}$ are arbitrary constant vectors, represents an ellipse. Express the equation of the ellipse in Cartesian coordinates. If $\mathbf{a} \cdot \mathbf{b} = 0$, show that $\mathbf{a}$ and $\mathbf{b}$ are the principal axes of the ellipse.

6. Transformation Properties of Vectors.—When a rectangular system of coordinates is used, a vector can be completely specified by its components. These components depend, of course, upon the orientation of the coordinate system, and the same vector may be described by many different triplets of components, each of which refers to a particular system of axes. The three components that represent a vector in one set of axes will be related to the components along another set of axes as are the coordinates of a point in the two systems. In fact, the components of a vector may be regarded as the coordinates of the end of the vector drawn from the origin. This fact is expressed by saying that the scalar components of a vector transform as do the coordinates of a point.

It is possible to concentrate attention entirely on the three components of a vector and to ignore its geometrical aspect. A vector would then be defined as a set of three numbers that

transform as do the coordinates of a point, when the system of axes is rotated. It is often convenient to designate the coordinate axes by numbers instead of the letters x, y, z so that the components of a vector will be a_1, a_2, and a_3. The designation for the whole vector is then a_i where it is understood that the subscript i can take on the values 1, 2, or 3. A vector equation is then written

$$a_i = b_i \tag{8-15}$$

This represents three equations, one for each value of the subscript i.

The rotation of a system of coordinates about the origin may be represented by the nine quantities $\gamma_{ij'}$, where $\gamma_{ij'}$ is the cosine of the angle between the i axis in one position of the coordinates and the j axis in the other position. These nine quantities give the angles made by each of the axes in one position with each of the axes in the other. They are also the coefficients in the expression for the transformation of the coordinates of a point. The cosines can be conveniently kept in order by writing them in the form of a matrix

$$\begin{matrix} \gamma_{11'} & \gamma_{12'} & \gamma_{13'} \\ \gamma_{21'} & \gamma_{22'} & \gamma_{23'} \\ \gamma_{31'} & \gamma_{32'} & \gamma_{33'} \end{matrix} \tag{8-16}$$

Of the nine quantities only three are independent since there are six independent relations between them. Since $\gamma_{ij'}$ can be considered as the component along the j' axis in one coordinate system of a unit vector along the i axis in the other,

$$\gamma_{i1'}^2 + \gamma_{i2'}^2 + \gamma_{i3'}^2 = \sum_{j'} \gamma_{ij'}^2 = 1 \tag{8-17}$$

This will be true for every value of i. Similarly,

$$\sum_{i} \gamma_{ij'}^2 = 1 \tag{8-17a}$$

The components of a vector, or the coordinates of a point, can be transformed from one system of coordinates to the other by

$$a_i = \gamma_{i1'}a_{1'} + \gamma_{i2'}a_{2'} + \gamma_{i3'}a_{3'} = \gamma_{ij'}a_{j'} \tag{8-18}$$

The $a_{j'}$ represents the components of the vector **a** in one system of coordinates and the a_i the components in the other. The summation sign is omitted in the last term since it is to be understood that a sum is to be carried out over all three values of any index that is repeated.

Problem 27. Show that

$$\gamma_{ij'}\,\gamma_{kj'} = \delta_{ik} \tag{8-17b}$$

where $\delta_{ik} = 1$ when $i = k$ and $\delta_{ik} = 0$ when $i \neq k$.

Problem 28. Write the components of the transformation matrix when one system of coordinates differs from the other by a rotation through the angle α about one axis.

Problem 29. Write the components of the transformation matrix when one system of axes is obtained from the other by rotation through 120° about an axis making equal angles with the three coordinate axes.

7. Linear Vector Functions.—If a vector is a function of a single scalar variable, such as the time, each component of the vector is independently a function of this variable. If the vector is a linear function of the time, each component is proportional to the time.

A vector may also be a function of another vector. In general, this implies that each component of the function depends on each component of the independent vector. A vector is a linear function of another vector if each component of the first is a linear function of the three components of the second. This requires nine independent coefficients of proportionality. The statement that **a** is a linear function of **b** means that

$$\left.\begin{aligned} a_1 &= C_{11}b_1 + C_{12}b_2 + C_{13}b_3 \\ a_2 &= C_{21}b_1 + C_{22}b_2 + C_{23}b_3 \\ a_3 &= C_{31}b_1 + C_{32}b_2 + C_{33}b_3 \end{aligned}\right\} \tag{8-19}$$

Using the summation convention as in equation (8-18), this becomes

$$a_i = C_{ij}b_j \tag{8-19a}$$

A relationship such as that in equation (8-19a) must be independent of the coordinate system in spite of the fact that the notation is clearly based on specific coordinates. The

components a_i and b_i are with reference to a particular coordinate system. The constants C_{ij} also have reference to specific axes, but they must so transform with a rotation of axes that a given vector **b** always leads to the same vector **a.**

If the coordinate system is rotated about the origin, the vector components will change so that

$$a_i = \gamma_{ij'}a_{j'} = C_{ij}\gamma_{jk'}b_{k'} \tag{8-20}$$

If both sides of this equation are multiplied by $\gamma_{l'i}$ and the equations for the three values of i are added, the result is

$$\gamma_{l'i}\gamma_{ij'}a_{j'} = a_{l'} = (\gamma_{l'i}C_{ij}\gamma_{jk'})b_{k'} \tag{8-21}$$

If the quantity $\gamma_{l'i}C_{ij}\gamma_{jk'}$ is called $C_{l'k'}$

$$a_{l'} = C_{l'k'}b_{k'} \tag{8-22}$$

This relationship between the components in this system of coordinates is the same vector relationship as was expressed by the C_{ik} in the original system of coordinates.

8. Tensors.—*Tensor* is a general name given to quantities that transform in prescribed ways when the coordinate system is rotated. A scalar is a tensor of rank 0, for it is independent of the coordinate system. A vector is a tensor of rank 1. Its components transform as do the coordinates of a point. A tensor of rank 2 has components that transform as do the quantities C_{ij}.

Tensors can be added or subtracted by adding or subtracting their corresponding components. They can also be multiplied in various ways by multiplying components in various combinations. These and other possible operations with tensors will not be described here. They can be found in some of the references suggested below.

A tensor of the second rank is said to be symmetric if $C_{ij} = C_{ji}$ and to be antisymmetric if $C_{ij} = -C_{ji}$. An antisymmetric tensor has its diagonal components equal to zero. Any tensor may be regarded as the sum of a symmetric and an antisymmetric part for

$$C_{ij} = \tfrac{1}{2}(C_{ij} + C_{ji}) + \tfrac{1}{2}(C_{ij} - C_{ji}) \tag{8-23}$$

and

$$\tfrac{1}{2}(C_{ij} + C_{ji}) = S_{ij} \qquad \tfrac{1}{2}(C_{ij} - C_{ji}) = A_{ij}$$

where S_{ij} is symmetric and A_{ij} is antisymmetric.

Numerous physical quantities have the properties of tensors of the second rank. As will be shown in the next chapter, the inertial properties of a rigid body can be described by the symmetric tensor of inertia.

Problem 30. A particle is attracted toward each of the three coordinate planes with a force proportional to the distance but with different proportionality constants in the three directions. The total force is then a linear vector function of the displacement. Write the tensor relating the force and the displacement. Show how it is transformed when the axes are rotated about one of them.

Problem 31. Show that if C_{ij} is a symmetric tensor and

$$a_i C_{ij} a_j = 1 \tag{8-24}$$

the vectors that satisfy this equation reach from the origin to a quadratic surface. Show that, if the diagonal components of the tensor are all positive and the nondiagonal components all zero, the surface is an ellipsoid.

Problem 32. Show that $C_{ij}D_{jk} = E_{ik}$, where E_{ik} is a tensor of the second rank, *i.e.*, it transforms as does C_{ij}.

References

COFFIN, J. G.: "Vector Analysis," John Wiley & Sons, Inc., New York.

GIBBS, J. WILLARD, and E. B. WILSON: "Vector Analysis," Charles Scribner's Sons, New York.

PAGE, L.: "Introduction to Theoretical Physics," Chap. 1, D. Van Nostrand Company, Inc., New York.

CHAPTER IX

DYNAMICS OF RIGID BODIES

The motion of a solid body can be treated by applying to each small part of it the laws applicable to the motion of a particle. These particles will move under the influence of forces from other parts of the body, as well as outside forces, and both types of forces must be taken into account. The application of this analysis to every particular case would be extremely complicated, but it has been found possible to define a number of idealizations that approximate practically important cases. These idealizations permit the establishment of certain simplifying theorems, so that the motion of the body under consideration becomes amenable to a mathematical treatment.

One such idealization is the elastic body, with relationships between internal stresses and deformations that conform to Hooke's law. Another idealization, and the one that will be treated in this chapter, is the *rigid body*. In a rigid body the distance between each pair of points is a constant and is never changed by the application of a force.

Strictly speaking, there exist in nature no rigid bodies. All bodies can be deformed by the application of sufficient force, and in many cases this deformation has an important influence on the motion. Nevertheless, the ideal rigid body is a useful approximation. It is possible to prove a number of theorems and to define a number of quantities that refer to such a body as a whole, rather than to its elementary parts. These theorems constitute the laws of motion of rigid bodies. It should be emphasized, however, that such theorems are merely consequences of Newton's laws and the assumed rigidity of the bodies. They contain no really new statements.

1. Center of Mass and Linear Momentum.—The center of mass of a rigid body can be defined in a manner similar to that

used for the center of mass of a number of particles. Each volume element of the body is considered as a particle, and the sums are replaced by integrals.

$$\mathbf{R} = \frac{1}{M} \int \mathbf{r}\rho \, dv \tag{9-1}$$

where $\mathbf{R}$ is the vector from some origin to the center of mass and $\mathbf{r}$ is the vector from the same origin to the volume element dv. ρ is the mass density at the volume element, and M is the total mass of the body. The integral is taken over the whole of the body in question.

The properties of the center of mass of a rigid body are essentially the same as those of the center of mass of a collection of particles. Indeed, there is some simplification in the case of the rigid body, since the distances between the various parts of the body are constant and the center of mass has a fixed location in it.

Problem 1. Find the location of the center of mass of an octant of a sphere of uniform density.

Problem 2. Find the center of mass of a rectangular parallelepiped whose density increases linearly from one end to the other.

When Newton's equations of motion are applied to each element of a rigid body, an equation for the motion of the center of mass can be obtained.

$$\int \rho \frac{d^2\mathbf{r}}{dt^2} \, dv = M \frac{d^2\mathbf{R}}{dt^2} = \int {}_v\mathbf{F}_i \, dv + \int {}_v\mathbf{F}_e \, dv \tag{9-2}$$

The first equality follows from equation (9-1), and the second is just the integral of the forces acting on each volume element. ${}_v\mathbf{F}_i$ is the force per unit volume acting on the element dv due to the rest of the rigid body. ${}_v\mathbf{F}_i \, dv$ is then the total internal force on this volume element, and since the internal forces follow Newton's third law of motion their integral is zero. The vanishing of this integral comes about in just the same way as the vanishing of the sum of mutual forces on page 25.

In the same way ${}_v\mathbf{F}_e$ represents the vector external force per unit volume on the element of volume dv. The force per unit

volume is used here so that the product ${}_v\mathbf{F}_e\, dv$ will be the vector force on dv. If the force is a gravitational force, it has the magnitude ρg per unit volume. If the external forces are discrete forces, the integral is just the sum of the forces.

The vector linear momentum $\mathbf{P}$ of a rigid body is an integral over the vector momenta of the various volume elements so that

$$\int \rho \frac{d\mathbf{r}}{dt}\, dv = M \frac{d\mathbf{R}}{dt} = \mathbf{P} \tag{9-3}$$

It follows then from equation (9-2) that

$$\frac{d\mathbf{P}}{dt} = \int {}_v\mathbf{F}_e\, dv = \mathbf{F} \tag{9-4}$$

where $\mathbf{F}$ is the total external force acting on the body.

It is important to notice the significance of the integral in equation (9-4). The force effective in producing an acceleration of the center of mass is the sum of all the forces acting on all the particles of the body. It makes no difference at what points they are applied. If all of the force is applied at one or more discrete points, equation (9-4) states that the motion of the center of mass is determined by the vector sum of these forces but is independent of their points of application. This does not mean that the motions of all the parts of the body are independent of the points of application of the forces. It refers merely to the motion of the center of mass.

2. Angular Momentum.—The equation of motion for an element of volume dv is

$$\rho \frac{d^2\mathbf{r}}{dt^2}\, dv = {}_v\mathbf{F}_i\, dv + {}_v\mathbf{F}_e\, dv \tag{9-5}$$

as was used in equation (9-2). If the vector product of the left-hand side by $\mathbf{r}$ is integrated over the body, the result is the rate of change of the total angular momentum.

$$\int \rho\mathbf{r} \times \frac{d^2\mathbf{r}}{dt^2}\, dv = \frac{d}{dt} \int \rho\mathbf{r} \times \frac{d\mathbf{r}}{dt}\, dv = \frac{d\mathbf{H}_0}{dt} \tag{9-5a}$$

where $\mathbf{H}_0 = \int[\rho\mathbf{r} \times (d\mathbf{r}/dt)]dv$. The integral is taken over the entire body, and the quantity $d\mathbf{r}/dt$ is the vector velocity of the volume element dv.

The differentiation with respect to time can be taken out of the integral sign because of the assumed rigidity of the body. This involves the constancy, with respect to time, of ρ as well as of the dimensions of the body over which the integration is carried out. In these integrals the vector $\mathbf{r}$ is measured from a fixed origin, which may or may not be in the body. The subscript on $\mathbf{H}_0$ indicates that there is a fixed origin from which the vectors start.

The vector product of $\mathbf{r}$ with the right-hand side of equation (9-5) leads to two integrals. The first is

$$\int \mathbf{r} \times {}_v\mathbf{F}_i \, dv = 0 \tag{9-5b}$$

The value zero for this integral may be taken as part of the definition of a rigid body. It would follow from the assumption that the forces between the parts of a body are central forces, but this assumption lacks generality and seems unnecessarily restrictive. Frequently equation (9-5*b*) is justified by the statement that the internal forces within a rigid body are in equilibrium among themselves. It may be considered as a law similar to Newton's third law of motion. It describes the observed fact that isolated bodies do not set themselves into rotation.

The second integral is given a name.

$$\int \mathbf{r} \times {}_v\mathbf{F}_e \, dv = \mathbf{L}_0 \tag{9-5c}$$

where $\mathbf{L}_0$ is called the *torque* or *moment of force*, with reference to the origin about which it is measured.

It was pointed out above that the effect of forces on the motion of the center of mass is entirely independent of their points of application. However, the whole effect of the forces is not given by the motion of the center of mass. The rotation of the body does depend on these points of application through the law connecting $d\mathbf{H}_0/dt$ and $\mathbf{L}_0$. The additional law based on the idealization of a rigid body is the vector equation

$$\frac{d\mathbf{H}_0}{dt} = \mathbf{L}_0 \tag{9-6}$$

The remainder of this chapter is essentially a discussion of the consequences of this law.

Problem 3. Write out the Cartesian components of $\mathbf{H}_0$.

Problem 4. Write out the Cartesian components of $\mathbf{L}_0$.

Problem 5. Consider a uniform rod pivoted at its upper end so as to swing freely under gravity in a vertical plane. Evaluate the torque on it from equation (9-5*c*) and the angular momentum from equation (9-5*a*).

Problem 6. Show that the torque about the center of mass due to the weights of the elementary parts of a rigid body is zero.

3. Moments of Force and Couples.—Equation (9-2) shows that the motion of the center of mass of a rigid body is determined by the vector sum of all the acting forces but is independent of their various points of application. The system of forces could be replaced, as far as the translation of the center of mass is concerned, by a single force equal to the resultant and applied at any point whatever.

Equation (9-6) shows that the rate of change of the angular momentum of a rigid body depends on the total moment of force, or the total torque, but is independent of the total force. This torque is defined in terms of the forces and their points of application and is with reference to a certain origin. It is, however, a vector, so that it is independent of the orientation of the coordinate system in terms of which it is described.

To predict the motion of every point in a rigid body, it is both necessary and sufficient to know the resultant force acting and the total moment of force. These two quantities are quite independent. It is possible to have a set of forces whose resultant is zero but which produce a nonzero torque. In such a case the center of mass will move in a straight line with uniform velocity, and the rotation may conveniently be referred to it as a center.

The simplest combination of forces that has a moment but a zero resultant is that of two equal and parallel but oppositely directed forces that do not act in the same straight line. This combination is called a *couple*. The moment of a couple is independent of the origin to which it is referred. The vector

that represents it is perpendicular to the plane containing the two forces and has the sense of the angular momentum it tends to produce.

In their effects on a rigid body, two sets of forces are equivalent if they produce the same motion of the center of mass and the same kind of rotation. This requires that they have the same resultant and the same torque with reference to all origins. It can be shown, however, that if two sets of forces have the same resultant, and the same moment about one point, they have the same moment about all points. Hence it is necessary to require only the equality of the moments about one point. In general, a set of forces is not equivalent to a single resultant

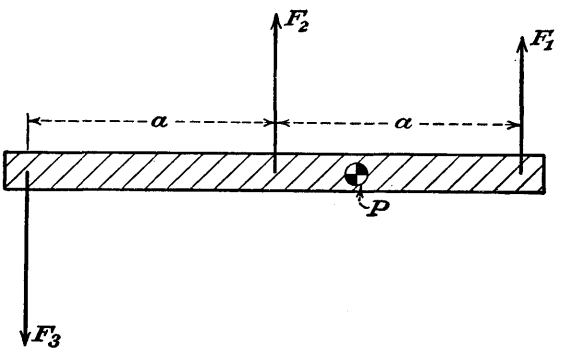


FIG. 9-1.—Three forces applied to a rigid rod.

alone. Such a set is equivalent, however, to a resultant applied at an arbitrarily selected point and a couple that depends upon the point selected for application of the resultant. It is necessary only that the couple be equal to the total moment of the set of forces about the point of application chosen for the resultant.

Problem 7. Show that the torque of a couple with reference to any origin is given by

$$\mathbf{L} = \mathbf{d} \times \mathbf{F} \tag{9-7}$$

where $\mathbf{F}$ is one of the two parallel forces and $\mathbf{d}$ is the vector to the point of application of this force from that of the other.

Problem 8. Consider the set of forces shown in Fig. 9-1 acting on a rigid rod. All three forces lie in the same plane. Find the single force and couple, applied at the left end of the rod, equivalent to this

set of forces. Find also the force and couple, applied at the point P a distance $2a/3$ from the right end of the rod, equivalent to the set shown.

Problem 9. If in the system of Fig. 9-1 the force $\mathbf{F}_1$ is turned 90° so as to act outward from the plane of the figure, find the equivalent force and couple applied at the center of the rod.

Problem 10. Show that, if the resultant of a system of forces is zero, the total moment about one point is the same as about any other.

Problem 11. Show that a single force applied at a given point of a rigid body is equivalent to the same force applied at any other point, plus a suitable couple.

Problem 12. Show that any set of forces acting on a rigid body can be replaced by an equivalent single force, acting at any point, and a suitable couple.

4. Kinematics of a Rigid Body.—To describe completely the position of every point in a body, it is both necessary and sufficient to give the positions of three selected points of the body that are not in the same straight line. Specifying the positions of three points requires, in general, nine coordinates; but since the condition of rigidity fixes the three distances between the points, only six independent coordinates are necessary to locate the points when they are in a rigid body. When the location of these three points is given, the location of every other point in the body is fixed by its distances from these three.

It is usually convenient to take three coordinates to locate one of the selected points. This point can then be taken as the origin of a system of coordinates fixed in the body. The line connecting this point with one of the other selected points can then be taken as the x' axis of this system, and two coordinates, ordinarily two angles, will give the direction in space of this axis. If the plane containing the x' axis and the third selected point is taken as the x'-y' plane, a single angle will give the orientation of this plane. The z' axis is then fixed by its relationship to the x'-y' plane. Hence three coordinates will locate the origin of the x'-y'-z' system, and three angles will give its orientation. Every point in the body has a fixed location in this system so that the six coordinates suffice for a complete specification of the location of every point.

The position of a particle can be specified by a vector from some fixed origin to the particle, and the most general displacement can be described in terms of a vector from an initial position to a final position. To describe the displacement of a rigid body is more complicated. It can be shown that the most general displacement can be described in terms of the translation of some particular point in the body and a rotation about an axis passing through this point. By such a description the displacement of every point in the body can be specified. However, the details of this process will not be treated here since they are unnecessary for the elementary dynamics of rigid bodies.

For dynamics it is more important to be able to describe the velocities and accelerations of each element of the body than to describe the displacements. For this purpose Chasles' theorem states that, in the most general case, the velocities of the parts of a rigid body can be described in terms of a translational velocity of any selected point in it, plus an angular velocity about an axis through this point.

The velocity of translation, and both the magnitude and direction of the angular velocity, will be functions of the time, but at any fixed time the angular velocity will be independent of the point about which it is measured.

Chasles' theorem can be established by proving it with reference to a coordinate system x', y', z', fixed in the body, since the motion of such a coordinate system fixes the motion of the body. The origin of this coordinate system is put at the point through which it is desired to have the axis of the angular velocity pass. Then let $\mathbf{a}$ be the vector from a fixed origin to the origin of the moving coordinates, and let $\mathbf{r}$ be the vector from the same fixed origin to an arbitrary point in the body. Let the coordinates of this point be (x',y',z') with reference to the system fixed in the body, so that x', y', and z' are constants, independent of the time. The vector $\mathbf{r}'$ from the origin of the moving coordinates to the arbitrary point has the constant components x', y', and z'. Then

$$\mathbf{r} = \mathbf{a} + \mathbf{r}'$$

and

$$\frac{d\mathbf{r}}{dt} = \frac{d\mathbf{a}}{dt} + \frac{d\mathbf{r}'}{dt} \tag{9-8}$$

$d\mathbf{a}/dt$ is the translational velocity of the origin and may be represented by $\mathbf{v}$; it remains to be shown that $d\mathbf{r}'/dt$ can be expressed in terms of an angular velocity.

The vector $\mathbf{r}'$ is fixed with reference to the moving system of coordinates, since x', y', z' are constants. By the derivative of $\mathbf{r}'$ with respect to the time is meant its rate of change with reference to a coordinate system whose axes are fixed in direction but whose origin coincides with the origin of the coordinates fixed in the moving body. Since the length of $\mathbf{r}'$ is constant, its only change is in direction and can be expressed in terms of the motion of the coordinate axes. Under these conditions

$$\frac{d\mathbf{r}'}{dt} = x'\frac{d\mathbf{i}'}{dt} + y'\frac{d\mathbf{j}'}{dt} + z'\frac{d\mathbf{k}'}{dt} \tag{9-9}$$

The unit vectors are vectors of constant length, and their rates of change must consequently be perpendicular to the vectors themselves. Hence let

$$\frac{d\mathbf{i}'}{dt} = c\mathbf{j}' + f\mathbf{k}' \qquad \frac{d\mathbf{j}'}{dt} = e\mathbf{i}' + a\mathbf{k}' \qquad \frac{d\mathbf{k}'}{dt} = b\mathbf{i}' + d\mathbf{j}' \tag{9-10}$$

Since the vectors $\mathbf{i}'$, $\mathbf{j}'$, and $\mathbf{k}'$ remain perpendicular to each other during the motion, $\mathbf{i}' \cdot \mathbf{j}'$, $\mathbf{j}' \cdot \mathbf{k}'$, and $\mathbf{k}' \cdot \mathbf{i}'$ are constant and the derivatives of these scalar products must be zero. This leads, for example, to

$$\frac{d\mathbf{i}'}{dt} \cdot \mathbf{j}' + \mathbf{i}' \cdot \frac{d\mathbf{j}'}{dt} = c + e = 0$$

and to the three relationships

$$d = -a \qquad f = -b \qquad e = -c \tag{9-10a}$$

If a, b, and c are taken as the three components of a vector $\boldsymbol{\omega}$,

$$\boldsymbol{\omega} = a\mathbf{i}' + b\mathbf{j}' + c\mathbf{k}' = \omega_{x'}\mathbf{i}' + \omega_{y'}\mathbf{j}' + \omega_{z'}\mathbf{k}' \tag{9-10b}$$

the relationships (9-10) can be combined with (9-9) to give

$$\frac{d\mathbf{r}'}{dt} = \begin{vmatrix} \mathbf{i}' & \mathbf{j}' & \mathbf{k}' \\ \omega_{x'} & \omega_{y'} & \omega_{z'} \\ x' & y' & z' \end{vmatrix} = \boldsymbol{\omega} \times \mathbf{r}' \tag{9-11}$$

The vector $\boldsymbol{\omega}$ is called the angular velocity.

The angular velocity $\boldsymbol{\omega}$ is defined in terms of quantities that refer to the motion of the coordinate system only, and thus it is the same for all points of the body. The combination of equations (9-11) and (9-8) then shows that

$$\frac{d\mathbf{r}}{dt} = \mathbf{v} + \boldsymbol{\omega} \times \mathbf{r}' \tag{9-12}$$

Problem 13. Show that, for the vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ from the origin of a moving system of coordinates to two points fixed in the coordinate system,

$$\mathbf{r}_1 \cdot \frac{d\mathbf{r}_2}{dt} + \frac{d\mathbf{r}_1}{dt} \cdot \mathbf{r}_2 = 0$$

Show that the conditions (9-10*a*) constitute a special case of this equation.

Problem 14. A plane body is moving in its own plane, and its motion is specified by the translation of one point and an angular velocity about an axis perpendicular to the plane and passing through the point. Find the point in the body that is instantaneously at rest, and show that every other point has a velocity perpendicular to the line connecting it with this point.

Problem 15. Consider a wheel of a locomotive that is traveling at a constant speed. Describe the motion of the wheel in terms of the translation of the center and an angular velocity. Then describe it in terms of the translation of a point halfway between the center and the rim and the corresponding angular velocity.

Problem 16. Show that the angular velocity of a solid body is the same for all locations of the origin about which it is taken.

Problem 17. Show that the angular-velocity vector is parallel to the axis of rotation.

Problem 18. Show that if the vector $\mathbf{a}$ has the components $a_{x'}$, $a_{y'}$, $a_{z'}$ in a moving system of coordinates

$$\frac{d\mathbf{a}}{dt} = \frac{da_{x'}}{dt}\mathbf{i}' + \frac{da_{y'}}{dt}\mathbf{j}' + \frac{da_{z'}}{dt}\mathbf{k}' + \boldsymbol{\omega} \times \mathbf{a} \tag{9-13}$$

5. Three Systems of Coordinates Axes.—In dealing with the motions of rigid bodies it is desirable, at different times, to use

different sets of coordinate axes. Of these, three are of particular importance.

a. Axes Fixed in Space with Reference to Which All Processes May Be Described.—Although there is some theoretical difficulty in defining the term *fixed in space,* there is no practical difficulty when dealing with ordinary phenomena on a small scale. It is with reference to such coordinates that Newton's laws were stated, and it is with reference to such coordinates that equations (9-2) and (9-6) were written.

b. Axes Fixed in Direction but with an Origin Fixed in the Moving Body.—There are two cases in which the selection of the origin for such a system will be different.

(1) If one or more points of the body are fixed in space, the origin will be located at one of them. If the body is a top spinning on its point or a wheel mounted as a gyroscope so that only one point is fixed, the selection of the origin is unique. In the case of a body turning about a fixed axis, any point on the axis is suitable for the origin. Since the directions of the axes are fixed and the origin is fixed, these coordinates are equivalent to those of (*a*) with a particular selection of the origin.

(2) If no point of the body is fixed, the origin is placed at the center of mass.

With respect to coordinates fixed in direction, integrals such as those defining angular momentum or torque must be carried out between changing limits. The limits will be the surfaces of the body and these will, in general, be moving.

c. Axes Fixed in the Rigid Body.—As in the previous case, the origin may be located differently in two cases.

(1) If there is a fixed point, it is usually convenient to locate the origin at such a point.

(2) If the body is entirely free, it is usually most convenient to locate the origin at the center of mass.

For those cases in which the origin is fixed, equation (9-6) applies directly. The integrations involved in the definitions of $\mathbf{H}_0$ and $\mathbf{L}_0$ must be carried out over the volume of the rigid bodies concerned. For this purpose, it is convenient to use axes fixed in the body when these integrals are to be evaluated, for then the integrations are between fixed limits. In principle, the integra-

tions can also be carried out between moving limits, with respect to fixed coordinates, if it is so desired.

For cases in which the origin is moving, it is important to consider equation (9-6) a little further. As above, let $\mathbf{r}'$ be the vector locating a point in the moving coordinate system, and let $\mathbf{a}$ be the vector from the fixed origin to the moving origin. Then

$$\begin{aligned}\mathbf{H}_0 &= \int \rho \mathbf{r} \times \frac{d\mathbf{r}}{dt}\, dv \\ &= \int \rho \left(\mathbf{a} \times \frac{d\mathbf{r}}{dt} + \mathbf{r}' \times \frac{d\mathbf{a}}{dt} + \mathbf{r}' \times \frac{d\mathbf{r}'}{dt}\right) dv \end{aligned} \tag{9-14}$$

Since the origin in these cases is at the center of mass,

$$\int \rho \mathbf{r}' \times \frac{d\mathbf{a}}{dt}\, dv = \left(\int \rho \mathbf{r}'\, dv\right) \times \frac{d\mathbf{a}}{dt} = 0 \tag{9-14a}$$

The time derivative of the angular momentum then becomes

$$\begin{aligned}\frac{d\mathbf{H}_0}{dt} &= \frac{d}{dt} \int \rho \left(\mathbf{a} \times \frac{d\mathbf{r}}{dt} + \mathbf{r}' \times \frac{d\mathbf{r}'}{dt}\right) dv \\ &= \int \rho \mathbf{a} \times \frac{d^2\mathbf{r}}{dt^2}\, dv + \int \rho \frac{d\mathbf{a}}{dt} \times \frac{d\mathbf{r}}{dt}\, dv + \frac{d}{dt} \int \rho \mathbf{r}' \times \frac{d\mathbf{r}'}{dt}\, dv \\ &= \mathbf{a} \times \int \rho \frac{d^2\mathbf{r}}{dt^2}\, dv + \frac{d\mathbf{a}}{dt} \times \int \rho \frac{d\mathbf{r}}{dt}\, dv + \frac{d\mathbf{H}}{dt} \end{aligned} \tag{9-14b}$$

where

$$\mathbf{H} = \int \rho \mathbf{r}' \times \frac{d\mathbf{r}'}{dt}\, dv \tag{9-14c}$$

Justification of the differentiation under the integral sign becomes obvious when it is remembered that the integral is equivalent to a sum over all the elements of the body.

In the last expression in equation (9-14*b*), the first integral is a force, and the second integral is equal to $M\, d\mathbf{a}/dt$ because the origin of $\mathbf{r}'$ is located at $\mathbf{r} = \mathbf{a}$. Hence

$$\frac{d\mathbf{H}_0}{dt} = \mathbf{a} \times \int {}_v\mathbf{F}_e\, dv + \frac{d\mathbf{H}}{dt} \tag{9-14d}$$

Also, from equation (9-5*c*)

$$\begin{aligned}\mathbf{L}_0 &= \int \mathbf{a} \times {}_v\mathbf{F}_e\, dv + \int \mathbf{r}' \times {}_v\mathbf{F}_e\, dv \\ &= \mathbf{a} \times \int {}_v\mathbf{F}_e\, dv + \mathbf{L} \end{aligned} \tag{9-14e}$$

where $\mathbf{L} = \int \mathbf{r}' \times {}_v\mathbf{F}_e \, dv$ is the torque with respect to the moving axes. From equations (9-6), (9-14*d*), and (9-14*e*), it follows that

$$\frac{d\mathbf{H}}{dt} = \mathbf{L} \tag{9-15}$$

Equation (9-15) shows that the relationship between the torque and the rate of change of angular momentum is the same in all the coordinate systems mentioned above. It is the same for a fixed origin and for an origin at the center of mass.

6. The Tensor of Inertia.—Since the motion of all the parts of a rigid body, relative to an origin fixed in the body, can be expressed in terms of an angular velocity, the quantity $d\mathbf{r}/dt$ in the definition of angular momentum can be replaced by $\boldsymbol{\omega} \times \mathbf{r}$. This gives, for $\mathbf{H}$,

$$\mathbf{H} = \int \rho \mathbf{r} \times (\boldsymbol{\omega} \times \mathbf{r}) dv = \int \rho [r^2 \boldsymbol{\omega} - (\boldsymbol{\omega} \cdot \mathbf{r})\mathbf{r}] dv \tag{9-16}$$

where the primes have been omitted and $\mathbf{r}$ is understood to have its origin at the origin of any one of the three systems of coordinate axes that are described above. The x component of this equation is

$$H_x = \omega_x \int \rho(y^2 + z^2) dv - \omega_y \int \rho xy \, dv - \omega_z \int \rho xz \, dv$$

As shown in the previous section, this definition is to be used for either fixed or moving coordinates; but if the origin of the coordinates is moving, it is to be taken at the center of mass. If the coordinate axes are fixed in the body, the limits of integration are constant and the integrals in (9-16) can be carried out in terms of the vector $\boldsymbol{\omega}$ and the dimensions of the body. For axes not fixed in the body, the result of the integration must be the same, although it is not so easy to express.

According to equation (9-16) the angular momentum is a function of the angular velocity and constants of the body. In fact, it is a linear vector function of the angular velocity. The relation between $\mathbf{H}$ and $\boldsymbol{\omega}$ can be written in tensor form by using the tensor of inertia. This has the nine components

$$I = \begin{bmatrix} \int \rho(y^2 + z^2) dv & -\int \rho xy \, dv & -\int \rho zx \, dv \\ -\int \rho xy \, dv & \int \rho(x^2 + z^2) dv & -\int \rho yz \, dv \\ -\int \rho zx \, dv & -\int \rho yz \, dv & \int \rho(x^2 + y^2) dv \end{bmatrix} \tag{9-17}$$

In terms of this tensor of inertia, the components of the angular momentum can be written

$$H_i = I_{ij}\omega_j \tag{9-18}$$

The diagonal components of the tensor of inertia are called the moments of inertia, and the nondiagonal components are called the products of inertia. If the coordinates are such that the products of inertia vanish, the coordinate axes are said to correspond to the principal axes of the body. It is always possible to rotate the coordinates so that the products of inertia vanish; *i.e.*, principal axes exist for all bodies and for all possible locations of the origin. These principal axes are of importance because of the simplification they introduce into the equations of motion.

Consider a cube of uniform density, and take the origin of coordinates at one corner with the coordinate axes lying along edges of the cube. The integrals over the cube are easily expressed in these coordinates; and if the edge of the cube is b,

$$I_{xx} = \rho \int_0^b dz \int_0^b dy(y^2 + z^2) \int_0^b dx = \tfrac{2}{3}\rho b^5 = \tfrac{2}{3}Mb^2 \tag{9-19}$$

where M is the total mass of the cube. Similarly,

$$I_{xy} = -\rho \int_0^b x\,dx \int_0^b y\,dy \int_0^b dz = -\tfrac{1}{4}Mb^2 \tag{9-19a}$$

Because of the symmetry of this problem, all the diagonal terms are identical, and all the nondiagonal terms are identical.

$$I = \begin{bmatrix} \frac{2}{3}Mb^2 & -\frac{1}{4}Mb^2 & -\frac{1}{4}Mb^2 \\ -\frac{1}{4}Mb^2 & \frac{2}{3}Mb^2 & -\frac{1}{4}Mb^2 \\ -\frac{1}{4}Mb^2 & -\frac{1}{4}Mb^2 & \frac{2}{3}Mb^2 \end{bmatrix} \tag{9-19b}$$

Clearly the edges of the cube are not principal axes.

If the cube is rotating at a constant rate about one edge, say the z axis, the angular velocity will be constant in the rotating coordinates, $\boldsymbol{\omega} = \omega_z\mathbf{k}$. Then the vector angular momentum has three components, and

$$\mathbf{H} = Mb^2\omega_z(-\tfrac{1}{4}\mathbf{i} - \tfrac{1}{4}\mathbf{j} + \tfrac{2}{3}\mathbf{k}) \tag{9-20}$$

It is important to notice that in this case, and in general, the angular momentum is not parallel to the angular velocity.

Hence, although the fundamental law requires that, in the absence of external torques, the angular momentum shall be constant, the angular velocity will not necessarily also be constant.

Problem 19. Work out the components of the tensor of inertia for a thin rectangular sheet, when the coordinate axes pass through the center of the sheet and are parallel to the edges. Show that the products of inertia are zero and

$$I_{xx} = \tfrac{1}{3}Mb^2 \qquad I_{yy} = \tfrac{1}{3}Ma^2 \qquad I_{zz} = \tfrac{1}{3}M(a^2 + b^2)$$

where $2a$ and $2b$ are the dimensions of the rectangle in the x and y directions, respectively.

Problem 20. Transform the above tensor to coordinates whose origin is at the center but whose x axis is along a diagonal of the rectangle.

Problem 21. Show that the moment of inertia of a body about any axis is equal to its moment of inertia about a parallel axis through the center of mass, plus the product of the total mass by the square of the distance between the two axes.

Problem 22. Show that the tensor of inertia of a body with reference to any system of axes is equal to the tensor of inertia with respect to a parallel system of axes of which the origin is at the center of mass, plus the tensor of inertia with reference to the first system of axes which represents the concentration of the whole mass of the body at the center of mass.

Problem 23. Show that the moment of inertia of a thin sheet about an axis perpendicular to its plane is equal to the sum of its moments about any two perpendicular axes which lie in the plane and intersect the first axis.

Problem 24. Find the angular momentum of a thin rectangular sheet when it is rotating about one of its diagonals.

Problem 25. Find the components of the tensor of inertia of a uniform sphere when a point on the surface is taken as the origin of coordinates.

A rigid body can have a very complicated shape, but its dynamical properties are given by its mass and its tensor of inertia. Since the tensor of inertia is symmetrical, only six independent quantities are needed to describe it. These quantities can be represented by an ellipsoid of inertia. This

is an ellipsoid whose center is at the origin of coordinates and of such a size and shape that the moment of inertia around any axis is inversely proportional to the square of the distance from the center to the surface of the ellipsoid along the direction in question.

To show that such an ellipsoid exists, consider the vector **r** with components r_1, r_2, and r_3, and transform the tensor of inertia to coordinates of which one axis lies along **r**. The cosines of the angles between **r** and the original axes are r_1/r, r_2/r, and r_3/r. Then the moment of inertia about this axis is given by

$$I_{rr} = \frac{r_i I_{ij} r_j}{r^2} \tag{9-21}$$

If the length of the vector **r** is chosen so that $1/r^2 = I_{rr}$, the components of **r** will satisfy the equation of an ellipsoid

$$r_i I_{ij} r_j = 1 \tag{9-22}$$

A point on the surface of this ellipsoid has the coordinates (r_1,r_2,r_3), which are connected by the relation (9-22), and is a distance r from the center.

Problem 26. Transform the tensor of inertia for a cube as given above to its form when one axis is a body diagonal of the cube. Find the principal axes of the ellipsoid of inertia for a cube when the origin is at one corner.

7. Euler's Equations.—The fundamental law of the rotation of rigid bodies is contained in equation (9-15). For many purposes, however, it is desirable to write out the component equations instead of combining them all into one vector equation. When this is done, two distinct cases must be considered. If the axes are chosen so as to be fixed in direction, the rate of change of the angular momentum must include the rates of change of the various components of the tensor of inertia, as well as the rates of change of the components of the angular velocity. This leads to a very complicated set of equations that are rarely, if ever, used. On the other hand, if coordinate axes are chosen that are fixed in the body, the components of the tensor of inertia are constant and the expression for the rate

of change of the angular momentum contains only the rates of change of the components of the angular velocity together with the terms that represent the motion of the coordinate axes. These equations take a particularly convenient form when the coordinate axes coincide with the principal axes of the body. In this latter case

$$\mathbf{H} = \mathbf{i}I_{xx}\omega_x + \mathbf{j}I_{yy}\omega_y + \mathbf{k}I_{zz}\omega_z \tag{9-23}$$

and

$$\frac{d\mathbf{H}}{dt} = \mathbf{i}I_{xx}\dot{\omega}_x + \mathbf{j}I_{yy}\dot{\omega}_y + \mathbf{k}I_{zz}\dot{\omega}_z + I_{xx}\omega_x\frac{d\mathbf{i}}{dt} + I_{yy}\omega_y\frac{d\mathbf{j}}{dt} + I_{zz}\,\omega_z\frac{d\mathbf{k}}{dt} \tag{9-24}$$

Equations (9-10) and (9-11) show how the rates of change of the unit vectors can be written in terms of the components of the angular velocity. When this is done, the three components of the equations of motion can be written separately as three scalar equations known as *Euler's equations*. Euler's equations referred to the principal axes are

$$\left.\begin{aligned} I_{xx}\dot{\omega}_x + (I_{zz} - I_{yy})\omega_z\omega_y &= L_x \\ I_{yy}\dot{\omega}_y + (I_{xx} - I_{zz})\omega_x\omega_z &= L_y \\ I_{zz}\dot{\omega}_z + (I_{yy} - I_{xx})\omega_y\omega_x &= L_z \end{aligned}\right\} \tag{9-25}$$

L_x, L_y, and L_z are the components of the external torque about the moving x, y, and z axes, respectively. It must be remembered that these components of torque are referred to moving axes and hence may change with the time even when the vector torque is constant.

It is possible to write similar equations when the axes are not parallel to the principal axes, but they are much more complicated and are rarely used.

Problem 27. Derive Euler's equations referred to principal axes.

Problem 28. Write Euler's equations when the axes are not principal axes.

8. The Eulerian Angles.—As has been already indicated above, six independent coordinates are necessary to specify the position of a rigid body. This is best done by specifying

the location and orientation of a system of axes attached to the body. Three of the coordinates will locate the origin of the axes, and the remaining three can be used to fix the orientation. This orientation can be specified in various ways. The nine components of the tensor Γ could be used to represent the positions of the axes in the body relative to an external set of fixed axes. Since there are six independent relationships between the nine components of the tensor, there are really only three independent quantities in it. This method, however, is not very

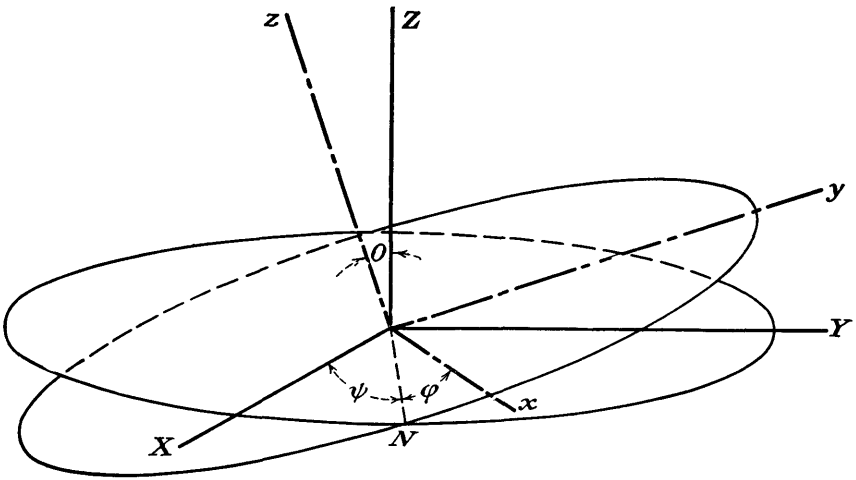


FIG. 9-2.—Specification of the Eulerian angles.

convenient. For many purposes, the most convenient coordinates are the Eulerian angles. They can serve to define the orientation of a rigid body, and they will also serve as independent variables in the equations of motion.

The Eulerian angles are indicated in Fig. 9-2. In this figure x, y, and z are the three axes fixed in the body, while X, Y, and Z represent a system of axes whose direction in space is fixed. The angle between the Z axis of the fixed system and the z axis of the moving system is called θ. The line ON in which the x-y plane cuts the X-Y plane is called the *line of nodes*, and the angle between the X axis and the line of nodes is called ψ. Some convention must be made as to the direction in which this angle is measured; therefore, it is usually specified to be measured in

the positive direction around OZ to the positive side of the line of nodes. The positive side of the line of nodes is defined as that side toward which the vector product of a vector along Z times a vector along z would point. The angle from the line of nodes, measured in the positive direction around z, to the positive x axis is called φ.

These three Eulerian angles seem an unsymmetrical set of coordinates, but they are especially adapted to the problem of a rotating body and give the equations the simplest form in most cases. Each of the angles can be changed independently of the others and the rate of change of each one can be considered as an angular velocity in the appropriate direction. By the projection of these angular-velocity vectors on the moving axes, it is possible to express the three Cartesian components of the angular velocity in terms of the derivatives of the Eulerian angles.

The angular velocity corresponding to θ can be seen from Fig. 9-2 to lie along the positive direction of the line of nodes. Its component along the x axis is then $\dot{\theta} \cos \varphi$. The angular velocity corresponding to $\dot{\varphi}$ is perpendicular to the plane including the line of nodes and the x axis and hence is along the z axis. This has no component along x. The angular velocity corresponding to $\dot{\psi}$ lies along the Z axis. This has a component $\dot{\psi} \cos \theta$ along Oz and a component $\dot{\psi} \sin \theta$ in the x-y plane. The latter component will also lie in the plane including z and Z. This plane is perpendicular to the line of nodes ON, and thus the component along x will be $\dot{\psi} \sin \theta \sin \varphi$. The total ω_x is the sum of these contributions. By similar analysis, the other angular velocities can be written down, and the result is

$$\left.\begin{aligned} \omega_x &= \dot{\theta} \cos \varphi + \dot{\psi} \sin \theta \sin \varphi \\ \omega_y &= -\dot{\theta} \sin \varphi + \dot{\psi} \sin \theta \cos \varphi \\ \omega_z &= \dot{\varphi} + \dot{\psi} \cos \theta \end{aligned}\right\} \tag{9-26}$$

Problem 29. Show from Euler's equations that if $I_{xx} = I_{yy}$ and $L_z = 0$, the angular velocity about the z axis is constant.

9. The Kinetic Energy of a Rotating Body.—The kinetic energy of a rigid body may be regarded as the sum of the kinetic

energies of its parts. Hence it can be written

$$T = \tfrac{1}{2}\int \rho(\dot{\mathbf{R}})^2\, dv = \int \tfrac{1}{2}\rho(\dot{\mathbf{a}} + \dot{\mathbf{r}}) \cdot (\dot{\mathbf{a}} + \dot{\mathbf{r}}) dv \qquad (9\text{-}27)$$

In this equation $\mathbf{R}$ is the vector from a fixed origin to the volume element dv, $\mathbf{a}$ is the vector from the fixed origin to the origin fixed in the rigid body, and $\mathbf{r}$ is the vector from the origin in the body to the volume element dv. According to the convention already described for the selection of systems of coordinates, $\mathbf{a}$ either is constant, or the origin of the moving system of coordinates is at the center of mass. For one reason or the other, equation (9-27) becomes

$$\begin{aligned} T &= \frac{M}{2}(\dot{a})^2 + \int \rho \dot{\mathbf{a}} \cdot \dot{\mathbf{r}}\, dv + \frac{1}{2}\int \rho \dot{\mathbf{r}} \cdot \dot{\mathbf{r}}\, dv \\ &= \frac{M}{2} v^2 + \frac{1}{2}\int \rho(\omega \times \mathbf{r}) \cdot (\omega \times \mathbf{r}) dv \end{aligned} \qquad (9\text{-}28)$$

In a triple vector product containing one dot and one cross, the dot and cross can be interchanged without affecting the value of the product. Hence

$$T = \frac{M}{2} v^2 + \frac{1}{2}\int \rho\omega \cdot \mathbf{r} \times (\omega \times \mathbf{r}) dv = \frac{1}{2} Mv^2 + \frac{1}{2}\omega \cdot \mathbf{H} \qquad (9\text{-}29)$$

Equation (9-29) contains the important result that the kinetic energy of a rigid body can be divided into two parts. One part is the energy of translation of the body as a whole, and the other part is the energy of rotation about the center of mass. Equation (9-29) shows that these two energies are entirely independent.

Since the angular momentum $\mathbf{H}$ is a linear vector function of the angular velocity, equation (9-29) shows that the kinetic energy of rotation is a homogeneous quadratic function of the velocity components.

Problem 30. Express the kinetic energy of a body that is turning about a principal axis, in terms of components of the tensor of inertia with reference to the principal axes.

Problem 31. Show that the kinetic energy of rotation of a rigid body rotating about any axis is equal to the moment of inertia about

the axis of rotation multiplied by one-half the square of the angular velocity.

Problem 32. Show from Euler's equations that the kinetic energy of rotation of a rigid body subject to no torque is constant.

Problem 33. Show that the rate of change of the kinetic energy of rotation of a rigid body is equal to the scalar product of the torque and of the angular velocity.

10. Rotation about a Fixed Axis.—The simplest motion of rotation is that in which one line in the body remains fixed and serves as an axis of rotation. Under these conditions the

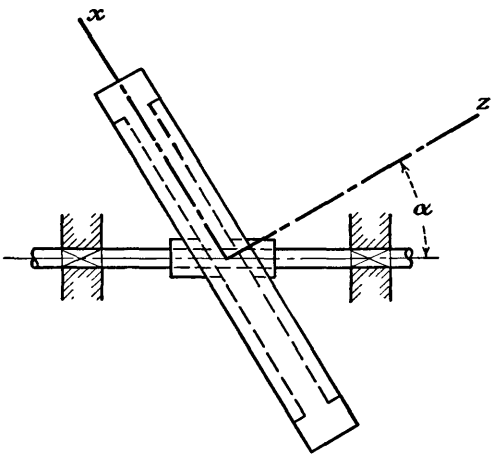


FIG. 9-3.—Moving axes used with a dynamically unbalanced flywheel.

angular-velocity vector is fixed in space and fixed in the body. However, the angular momentum is not always parallel to the angular velocity and thus the angular momentum is not fixed in direction, and a torque is needed to maintain such a motion. This torque is supplied through the bearings that support the axle. These bearings must supply the force necessary to cause the center of mass to move in its prescribed path, and at the same time they must supply the torque necessary to keep the axis of rotation fixed.

a. Flywheel out of Dynamic Balance.—Consider the case of a flywheel attached to its axle in such a way that, although the center of mass of the wheel is at the center of the axle, the principal axis of the system consisting of the wheel and axle is

not parallel to the axis of rotation but makes an angle α with it. Figure 9-3 shows the situation.

Since the center of mass is on the axis of rotation, it does not move when the wheel is turning and the only force that the bearings need supply to keep it in place is a constant upward force equal to the weight. This force w ll be divided between the two bearings according to their distances from the center of mass.

It is convenient to take axes fixed in the flywheel parallel to the principle axes of the body and with the origin at the center of mass. Take the z axis along the axis that makes the angle α with the axis of rotation. From the symmetry of the problem, it is apparent that another principal axis lies in the plane including z and the axis of rotation. Let the x axis be selected in this direction. The positive y axis will then be directed out of the paper. If α is small, the moments of inertia I_{xx} and I_{yy} will be practically equal. If they were exactly equal, the orientation of the x and y axes around the z axis would be arbitrary but it would still be convenient to select them as indicated.

The angular velocity is naturally along the axis of rotation so that if ω_0 is its magnitude

$$\boldsymbol{\omega} = -\omega_0 \sin\alpha\,\mathbf{i} + \omega_0 \cos\alpha\,\mathbf{k} \tag{9-30}$$

If I_{xx}, I_{yy}, and I_{zz} are the components of the tensor of inertia referred to the principal axes, the angular momentum in these coordinates is

$$\begin{aligned}\mathbf{H} &= \omega_x I_{xx}\mathbf{i} + \omega_z I_{zz}\mathbf{k} \\ &= \omega_0(-I_{xx}\sin\alpha\,\mathbf{i} + I_{zz}\cos\alpha\,\mathbf{k})\end{aligned} \tag{9-31}$$

All the quantities in this expression are constants except the unit vectors $\mathbf{i}$ and $\mathbf{k}$. But because the unit vectors are changing, $\mathbf{H}$ is changing, which shows the existence of a torque. From equation (9-10) it follows that

$$\frac{d\mathbf{i}}{dt} = \omega_z\mathbf{j} - \omega_y\mathbf{k} \qquad \text{and} \qquad \frac{d\mathbf{k}}{dt} = \omega_y\mathbf{i} - \omega_x\mathbf{j}$$

Hence

$$\mathbf{L} = \frac{d\mathbf{H}}{dt} = \omega_0^2(I_{zz} - I_{xx})\sin\alpha\cos\alpha\,\mathbf{j} \tag{9-32}$$

Since $I_{zz} > I_{xx}$ for an ordinary flywheel, this torque is in the positive direction around the y axis. But the y axis is turning uniformly in a plane perpendicular to the axis of rotation, and the force supplied by the bearings must change direction correspondingly. The force from one bearing, over and above the static force, will always be equal and opposite to that from the other since they constitute a couple about the center. The magnitude of each force will be constant, but the direction will rotate around the axis with the wheel. This tends to set the supporting structure into vibration and may cause undue stresses in the mounting

The wheel in this problem is in static balance. It will not tend to turn if it is set in any position. But it is not in dynamic balance and hence will require variable torques to keep it turning about a fixed direction. For smooth running it is just as important to have dynamic balance as to have static balance.

b. Case with Center of Mass off the Axis.—In the above illustration the center of mass was on the axis of rotation, and no force was needed to move it in the specified orbit. It often happens, however, that the center of mass is not on the axis of rotation, and this fact must be taken into account in computing the forces that must be supplied by the bearings.

As a simple illustration, consider a plane square sheet rotating about one edge and supported by bearings at the corners. For simplicity the effect of the weight (not the mass) will be neglected. For high speeds of rotation this may be quite justifiable. Figure 9-4 illustrates the situation and the axes selected with the origin at one bearing. The sheet is taken to be $2a$ on an edge and to be rotating about the x axis with a constant angular velocity ω_0.

The center of mass moves in a circle about the axis, but to make it do so a force toward the axis must be provided. Hence

$$\mathbf{F}_1 + \mathbf{F}_2 = -Ma\omega_0^2\mathbf{j} \tag{9-33}$$

The division of this force between the two bearings must be determined from the torque, and this is given by the rate of change of angular momentum. With the coordinates indicated

$$I_{xx} = \tfrac{4}{3}Ma^2 \qquad \text{and} \qquad I_{xy} = -Ma^2 \tag{9-34}$$

Hence

$$\mathbf{H} = Ma^2\omega_0(\tfrac{4}{3}\mathbf{i} - \mathbf{j}) \tag{9-35}$$

and

$$\mathbf{L} = \frac{d\mathbf{H}}{dt} = -Ma^2\omega_0{}^2\mathbf{k} \tag{9-36}$$

The force $\mathbf{F}_1$ produces no torque about the origin, and therefore all the torque must be due to $\mathbf{F}_2$ applied at the point $2a\mathbf{i}$.

$$2a\mathbf{i} \times \mathbf{F}_2 = \mathbf{L} = -Ma^2\omega_0{}^2\mathbf{k} \tag{9-37}$$

Equations (9-33) and (9-37) must be solved simultaneously to evaluate $\mathbf{F}_1$ and $\mathbf{F}_2$.

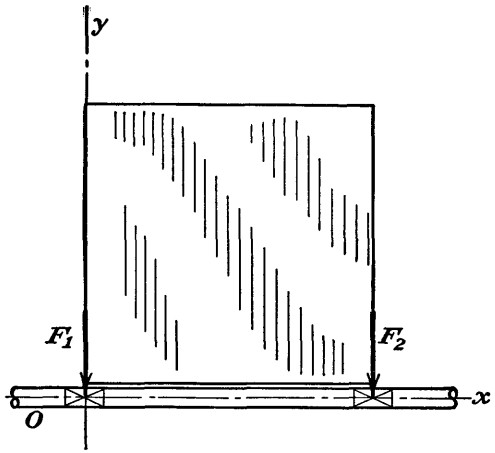


FIG. 9-4.—A plane sheet turning about a fixed axis.

The vector product of $\mathbf{i}$ and (9-33) gives

$$\mathbf{i} \times \mathbf{F}_1 + \mathbf{i} \times \mathbf{F}_2 = -Ma\omega_0{}^2\mathbf{k} \tag{9-33a}$$

Then (9-37) can be written

$$\mathbf{i} \times \mathbf{F}_2 = -\frac{M}{2}a\omega_0{}^2\mathbf{k} \tag{9-37a}$$

This leads to

$$\mathbf{i} \times \mathbf{F}_1 = -\frac{M}{2}a\omega_0{}^2\mathbf{k} \tag{9-38}$$

From this equation nothing can be concluded about the x component of $\mathbf{F}_1$, but the z component can be seen to be zero and the y component equal to $-\frac{1}{2}Ma\omega_0{}^2$.

Equation (9-33) shows that the sum of the x components of $\mathbf{F}_1$ and $\mathbf{F}_2$ is zero, but the individual components are not further determined. The motion is independent of what they may be. Hence assume them to be zero. The result is then

$$\mathbf{F}_1 = \mathbf{F}_2 = -\tfrac{1}{2}Ma\omega_0^2\mathbf{j} \tag{9-39}$$

Since the coordinate system is rotating, the direction of $\mathbf{j}$ is changing and the forces must follow it.

Problem 34. Evaluate the angular momentum of the square sheet rotating about one edge by the use of the integral definition (9-14*c*).

Problem 35. Treat the problem of the rotating square sheet when the origin of the moving coordinates is taken at the center of mass.

Problem 36. A body is suspended from a horizontal axis parallel to one of the principal axes that passes through the center of mass. The axis of rotation is a distance d from the center of mass. Write the Lagrangian function and the equations of motion for this type of pendulum.

11. Free Rotation of a Rigid Body.—If a body is subject to no external force whatever, the center of mass will move in a straight line with constant velocity and the angular momentum will be constant in both direction and magnitude. This, however, does not at all mean rotation with constant angular velocity about an axis fixed in direction. It was shown in the previous section that motion about a fixed axis may require a torque to maintain it. The motion of a free body with constant angular momentum may appear complicated if the body has an irregular shape, but it has certain simple characteristics that permit its description.

It was shown in Prob. 32 that the kinetic energy of a body subject to no forces is constant. This means that

$$2T = \boldsymbol{\omega} \cdot \mathbf{H} = \text{const} \tag{9-40}$$

and, combined with the constancy of $\mathbf{H}$, it means that the projection of $\boldsymbol{\omega}$ on $\mathbf{H}$ is constant. If $\boldsymbol{\omega}$ and $\mathbf{H}$ are drawn from a common origin, $\mathbf{H}$ will be constant in magnitude and direction while $\boldsymbol{\omega}$ will always end in a plane perpendicular to $\mathbf{H}$ at a distance $2T/H$ from the origin.

Problem 37. Show from the form of equation (9-40) that the vector $\boldsymbol{\omega}$ also always ends in the surface of an ellipsoid that is fixed in the body.

Problem 38. Show that, if the angular velocity of a free body coincides in direction with a principal axis, it will always coincide with that axis and will be fixed in space as well.

Problem 39. Show that the rotational kinetic energy of a body for which $I_{xx} = I_{yy}$, referred to principal axes, is

$$T = \tfrac{1}{2}I_{xx}(\dot{\theta}^2 + \dot{\psi}^2 \sin^2\theta) + \tfrac{1}{2}I_{zz}(\dot{\varphi}^2 + 2\dot{\varphi}\dot{\psi}\cos\theta + \dot{\psi}^2\cos^2\theta) \quad (9\text{-}41)$$

Problem 40. Show that for a rigid body, with $I_{xx} = I_{yy}$, the angular momentum vector lies in the plane containing the z axis and the angular velocity. If $I_{zz} > I_{xx}$, **H** lies between the z axis and $\boldsymbol{\omega}$. If $I_{zz} < I_{xx}$, the angular velocity lies between the z axis and **H**.

The study of the motion of a free rigid body is of importance in understanding the motion of projectiles shot from a rifle. These are projected with an angular velocity about the axis of symmetry; and, to the extent that air resistance can be neglected, such a projectile can be treated as a free body as far as its rotation is concerned. Gravity will accelerate the center of mass but will produce no torque or change in angular momentum. Since the treatment as a free body is based on the neglect of air resistance, questions of stability are important. It is important to know that small deviations from an ideal motion will not result in entirely new characteristics.

In Prob. 29 it was shown that the angular velocity about the z axis, which is taken as the axis of symmetry, is constant. Let this be called ω_0. When the projectile is fired, it is given this initial angular velocity about this principal axis; and, according to Prob. 38, it should continue to rotate about this axis, and the axis should remain fixed in direction. However, small torques may introduce small amounts of angular momentum in other directions, and it is important to see how this will affect the motion.

By differentiating the first of Euler's equations and substituting ω_y from the second it follows that

$$\ddot{\omega}_x + \left(\frac{I_{zz} - I_{xx}}{I_{xx}}\right)^2 \omega_0^2\omega_x = 0 \quad (9\text{-}42)$$

Then

$$\omega_x = A \sin \left[\left(\frac{I_{xx} - I_{zz}}{I_{xx}} \right) \omega_0 t - \epsilon \right] \qquad (9\text{-}42a)$$

where account is taken of the fact that $I_{xx} > I_{zz}$.

The solutions of Probs. 41 and 42 below permit one to show that the vector ω moves around the z axis so as to describe a circular cone. For bodies in which the difference between I_{xx} and I_{zz} is a significant fraction of I_{xx} the motion of this angular-velocity vector is almost as fast as the rotation of the body itself.

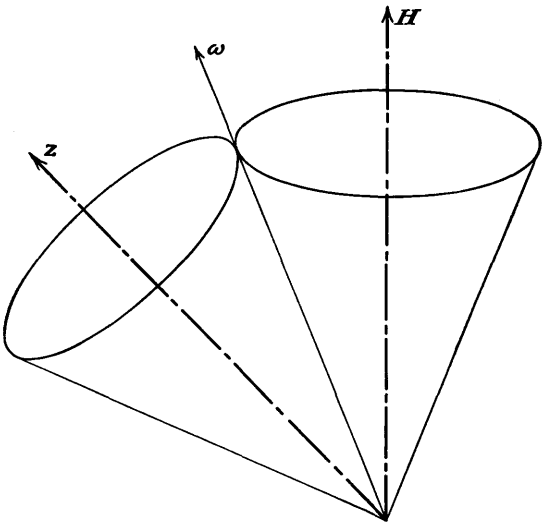


FIG. 9-5.—Representation of the motion of the principal axis of a free body z and the angular velocity vector ω by one cone rolling on another.

Problem 41. Show that for a symmetrical body

$$\omega_x^2 + \omega_y^2 = \text{const} \qquad (9\text{-}43)$$

Problem 42. Show from Euler's equations that ω_y follows a differential equation of the form of (9-42), and find the relationship between the constants in equation (9-42a) and the corresponding constants in the expression for ω_y.

Problem 43. A projectile consists of a right-circular cylinder with a hemispherical cap on one end. The length of the cylinder is three times its diameter. Find the rate at which the angular-velocity vector moves in the body compared with its rate of rotation.

As a consequence of the rotation of the body about an axis that is not a principal axis and the motion of the axis of rotation around the axis of symmetry, as well as the fact that the axis of symmetry, the angular velocity, and the angular momentum always lie in the same plane, the motion can be described as the rolling of one cone on another. Figure 9-5 indicates the situation. The vector **H** is composed of a portion $I_{zz}\omega_0$ along the z axis and a portion $I_{xx}(\omega_x^2 + \omega_y^2)^{\frac{1}{2}}$ perpendicular to z. The angular-velocity vector lies between z and **H** and in the same plane. If then a right-circular cone is constructed about **H** such that ω lies in its surface and another cone tangent to the first is constructed about z, the rigid body can be considered as attached to the second cone. The motion of the rigid body is then reproduced by rolling the second cone around the first.

12. Rotation of a Symmetrical Body about a Fixed Point.—Another case of interest is that in which a point in a symmetrical body is kept fixed in space and some external torques are applied. For simplicity take the fixed point to be the center of mass and let $I_{xx} = I_{yy}$. This is the case usually referred to as a gyroscope.

Problem 44. Use Lagrange's equation to obtain the equations of motion of a symmetrical body in terms of the Eulerian angles, and then show that, if the only torque acting is such that it tends to increase the angle θ, the following equations hold:

$$\left.\begin{aligned} I_{zz}(\dot{\varphi} + \dot{\psi}\cos\theta) &= R \\ I_{xx}\dot{\psi}\sin^2\theta + R\cos\theta &= S \\ I_{xx}\ddot{\theta} - I_{xx}\dot{\psi}^2\sin\theta\cos\theta + R\dot{\psi}\sin\theta &= L_\theta \end{aligned}\right\} \tag{9-44}$$

where R and S are constants of integration.

Problem 45. Show that $\theta = \pi/2$, $\dot{\varphi} = \text{const}$, $\dot{\psi} = \text{const}$ is a solution of equations (9-44) if there is a proper relationship between $\dot{\varphi}$ and $\dot{\psi}$.

Problem 46. Show that equations (9-44) admit of a solution corresponding to any desired constant angle of inclination θ_0 and any desired constant rate of precession $\dot{\psi} = \Omega$ if the torque L_θ has the proper constant value. Find the necessary value of the torque.

References

PAGE, L.: "Introduction to Theoretical Physics," Chap. II, D. Van Nostrand Company, Inc., New York.

SLATER and FRANK: "Introduction to Theoretical Physics," Chapter X, McGraw-Hill Book Company, Inc., New York.

CHAPTER X

THERMODYNAMICS

The subject of thermodynamics developed from the study of heat engines and concerns the transformation of heat into other kinds of energy. The work of Sadi Carnot, published in 1824, may be considered as the first real contribution to this study. Guided by the prevalent idea of his time, that heat was an indestructible substance, he sought for the conditions under which it would produce work. His conclusion was that heat can produce external work when it falls from one temperature to a lower temperature, much as water can do work in falling from one level to a lower level. Although this conclusion is only partly correct, Carnot's penetrating insight led to the discovery of the theorem that the maximum amount of work which can be obtained from a heat engine depends only upon the temperatures between which it operates and not at all on the working substance used.

By the middle of the nineteenth century the idea of heat as an indestructible substance had been superseded by the mechanical theory. According to this point of view, the increase in temperature of a body is the outward evidence of an increase in the energy of its molecules. The energy which is put into the body to produce this increase in temperature may be transferred by contact from another body at a higher temperature, or it may be transferred by doing mechanical work. Thus the idea of "quantity of heat" loses its definiteness and applies more to a method of transfer of energy than to any quantity which exists after the energy is transferred. The fundamental ideas of the mechanical theory of heat are very old. They were known to Huygens; and before Carnot's death, he himself recognized the insufficiency of the indestructible-substance idea with which he had been working. Between 1840 and 1850, however, Mayer, Helmholtz,

Joule, and others stated the conception so clearly, and performed the necessary experiments so convincingly, that this theory began to be widely accepted.

It was the combination of the mechanical theory of heat with the discoveries of Carnot about the availability of heat for doing work that produced the principles of thermodynamics. This synthesis was due to a large extent to Clausius and to William Thomson (Lord Kelvin). They put thermodynamics into its present position—that of a deductive science, founded on two very general laws. In their hands it became a powerful means of treating a restricted class of problems. More recent work has added a third general principle, which is now widely accepted as the third law of thermodynamics. This, however, will not be treated in the present chapter.

1. The Problems and Methods of Thermodynamics.

a. The Thermodynamic System.—In the treatment of thermodynamic problems, the objects under consideration are usually divided into two parts. One part is called the *system*, the other the *surroundings*. Attention is then given to the transfer of energy between the surroundings and the system by means of heat flow or by means of mechanical work done on the system. In the case of a steam engine the water may be considered as the system, while the rest of the engine constitutes the surroundings. Energy is given to the water by the heat flow from the fire, and some is returned to the condenser by the heat flow from the steam. In addition, an exchange of energy takes place between the steam and the piston. The study of the relationships between the energies transferred as heat and as work constitutes the problem of thermodynamics.

b. The State of a Thermodynamic System.—The state of a mechanical system is determined by the positions and the velocities of all its component parts. According to the mechanical theory of heat, the state of a thermodynamic system then would be determined by the positions and velocities of all of its component molecules. So detailed a knowledge of a system, however, is not necessary for the thermodynamic treatment of its behavior. It is necessary to know only the

quantities which are observable by macroscopic methods. These quantities are the pressure which the system exerts upon its container, the volume which it occupies, its temperature, its energy, the electric, magnetic, or gravitational fields to which it is subject, and any other parameters of a similar nature. These variables are not usually all independent, and the equations which connect them are known as the *equations of state.* The simplest cases are those in which the only variables involved in the determination of the state are the volume, pressure, and temperature. In such cases, two of the quantities are selected as the independent variables, and the third is expressed in terms of them by means of the equation of state. The energy, and any other functions of the state, can also be expressed in terms of the same two independent variables.

If the temperature T and the volume V of a given mass of a substance are taken as the independent variables of state, the pressure may then be written as a function of these two.

$$p = p(T,V) \tag{10-1}$$

The nature of this function can be determined by experiment only and is characteristic of the substance. For this reason the equation of state is sometimes called the *characteristic equation.* If the temperature and the volume are changed, the pressure is changed accordingly by an amount dp, which is given by

$$dp = \left(\frac{\partial p}{\partial T}\right)_V dT + \left(\frac{\partial p}{\partial V}\right)_T dV \tag{10-1a}$$

The subscripts attached to the partial derivatives indicate the variables that are kept constant during the differentiation. This method of specifying the independent variables is necessary because of the variety of such variables which can be used. Equation (10-1a) is also the equation of a curve of constant pressure, or an isobar, when dp is set equal to zero. In a similar fashion the equations for isothermal curves and curves of constant volume can be obtained.

Problem 1. Take the temperature and volume as independent variables, and draw curves representing states of equal volume,

equal pressure, and equal temperature. From these curves find the geometrical significance of the differential coefficients in equation (10-1*a*).

Problem 2. Take the pressure and the volume as independent variables. Draw for this case the curves indicated in the previous problem, and interpret the differential coefficients.

Problem 3. Repeat Prob. 2 for the case where temperature and pressure are the independent variables.

Problem 4. Show that

$$\left(\frac{\partial V}{\partial T}\right)_p = -\frac{(\partial p/\partial T)_V}{(\partial p/\partial V)_T} \tag{10-1b}$$

Problem 5. Show that

$$\left(\frac{\partial p}{\partial T}\right)_V = -\frac{(\partial V/\partial T)_p}{(\partial V/\partial p)_T} \tag{10-1c}$$

c. Thermometry.—The first ideas of heat and temperature come from the sensation of heat and cold, and from the observation that, if a hot body and a cold body are placed in contact, the hot body becomes less hot and the cold one becomes less cold. The distinction between *quantity of heat* and temperature becomes necessary when it is observed that a small hot body becomes nearly as cold as a large cold body with which it is placed in contact. It is also easy to observe a correlation between hotness and expansion, so that it is natural to use the expansion of some standard substance as a measure of temperature. Because of their large expansions, gases were early selected as the standard materials for thermometers. This, however, is a perfectly arbitrary definition of temperature, and it should not be expected that physical laws could be expressed in any simple form with this sort of temperature scale. The establishment of a natural scale of temperature was suggested by Lord Kelvin in 1848. This suggestion was based on the principle now known as the *second law of thermodynamics*, and its discussion will be left until later. It will suffice to say here that the absolute scale of temperature does not differ much from the scale of a gas thermometer.

The unit of heat is the calorie. This is the amount of heat required to raise the temperature of one gram of water by one

degree centigrade. Careful measurements show, however, that the initial temperature of the water makes some difference. If the above definition were satisfactory, the equilibrium temperature reached by mixing two equal masses of water at different temperatures would be the arithmetic mean of their initial temperatures. This is found not to be true, and therefore the calorie must be defined between two specific degrees of the centigrade scale. It is frequently defined between 15 and 16°C.

Experiments on the equilibrium temperature attained when other substances are mixed with water show that it is necessary to assign a number called the *heat capacity* to every body if it is desired to maintain the idea of a quantity of heat. Thus there arises the idea of specific heat. This is defined as the amount of heat necessary to raise the temperature of one gram of the substance through one degree. This might as well have been defined with reference to unit volume, but it is subject to less variation with external conditions if the definition with reference to unit mass is used. Further difficulties arise when a change of state is considered. The amount of heat necessary to raise one gram of water from 99.5 to 100.5°C at atmospheric pressure is certainly far from one calorie. To account for this fact, there was introduced the term *latent heat.* This term implies that when heat is used in changing the state of a substance it is not used in raising its temperature, and yet the heat, as such, is presumed to reside in the body.

All the terminology of thermometry is based upon the implicit idea of heat as an indestructible substance. With the aid of the various devices mentioned above, it is possible to maintain the fiction of a "quantity of heat," and to speak of the amount of heat in a body. With the complete acceptance of the mechanical theory, the idea of a quantity of heat must disappear; but the terms of the older theory can hardly be avoided, since they are in such common use. Strictly speaking, however, the only quantity which exists in this connection is energy. Energy may be transferred from a body at one temperature to another at a lower temperature by a heat flow. The

heat is only a means of transfer of energy and should not be imagined to exist inside the body.

For some time after the recognition of the mechanical nature of heat flow, attempts were being made to consider the kinetic energy of the molecules as heat, while the potential energy was regarded as internal energy. It is, however, of very doubtful advantage to try to separate these two kinds of energy; and throughout the rest of this chapter the term *heat* will be used only in connection with the flow of energy occasioned by a difference in temperature.

d. Reversible Processes.—If a gas is allowed to expand into a vacuum, it does no work on its surroundings, since there is no resisting force and the surroundings are not moved. If it expands against a resisting force, the work done is equal to the integral of the force multiplied by the distance moved. The most work is done when the resisting force is the largest. However, if the resisting force is larger than the pressure of the gas, there will be no expansion whatever, so that the maximum work will be done when the resisting force is just infinitesimally smaller than the force exerted by the gas. Under these circumstances the two forces are essentially equal to each other, and either can be used in calculating the work done. When expansion is carried out in this way, it is called *reversible*, because an infinitesimal increase in the opposing force will cause a compression.

If a heat flow occurs from one body to another at a lower temperature, the process is irreversible, since the energy cannot be made to flow directly from the colder to the warmer body. If, however, the temperature difference between the two bodies is infinitesimal, the rate of flow will be infinitely slow and an infinitesimal change in the temperature of one body will cause a heat flow in the opposite direction. The process of heat flow, then, is reversible when the two bodies are at practically the same temperature.

The reversibility of a process refers to the manner in which it is carried out rather than to the initial and the final states of the system. The first illustration given above shows that the

maximum amount of work is done on the surroundings when an expansion is carried out reversibly. It will be shown later that a reversible engine is the most efficient kind possible. For these reasons, analyses of thermodynamic problems are often made by reference to reversible processes.

e. Cyclical Processes.—It was first emphasized by Carnot that, if a process is to consist in the absorption of energy as heat and the loss of the same amount of energy as work, all the material bodies used in the process must be left at the end in exactly the state in which they were at the beginning. If this condition is fulfilled, the process will consist in one or more cycles of operations, such as expansions and compressions, absorptions and emissions of energy as heat, vaporizations and condensations, which return the system to its original state. In the course of such a cycle the system will, in general, absorb or give out a certain net amount of energy as heat and will do or receive a certain net amount of work. When a cyclical process is carried out in a reversible manner, the energy relations can be easily analyzed.

f. The Perfect Gas.—A perfect gas is one whose equation of state is

$$pV = nRT \tag{10-2}$$

where p is the pressure, V the volume of the gas, T the temperature measured on the absolute scale, R the so-called "gas constant," and n the number of gram molecules of the gas. $R = 8.314 \times 10^7$ ergs per degree per gram molecule.

The ordinary permanent gases are fair approximations to a perfect gas in the ordinary ranges of temperature and pressure, and reasoning carried out with equation (10-2) can be applied to them with a high degree of accuracy. It should always be remembered, however, that a perfect gas is a limiting case which is not actually realized in any real gas.

Problem 6. Show that, for a perfect gas,

$$\left(\frac{\partial p}{\partial V}\right)_T = \frac{-nRT}{V^2} \tag{10-2a}$$

Problem 7. Show that, for a perfect gas,

$$\left(\frac{\partial p}{\partial T}\right)_V = \frac{nR}{V} \tag{10-2b}$$

Problem 8. Take the temperature and volume as the independent variables, and draw the lines of constant pressure for a perfect gas.

Problem 9. Illustrate equations (10-1*b*) and (10-1*c*) by means of the equation of state for a perfect gas.

Problem 10. Show that the work done in compressing any substance from a volume V_1 to a volume V_2, by means of hydrostatic pressure p, is given by

$$W = -\int_{V_1}^{V_2} p\, dV \tag{10-3}$$

Problem 11. Show that the work done on a perfect gas in compressing it reversibly from a volume V_1 to the volume V_2 is

$$W = nRT \log \frac{V_1}{V_2} \tag{10-3a}$$

g. The van der Waals Equation of State.—Although the ordinary gases obey the perfect-gas law fairly well, vapors of high density and liquids deviate very far from it. An equation of state which describes vapors and liquids well enough for some purposes and which approaches the perfect-gas law as a limiting case is the van der Waals equation. This is

$$p = n\left(\frac{RT}{V - nb} - \frac{na}{V^2}\right) \tag{10-4}$$

The letters p, n, V, R, and T have the same significance as before. a and b are constants which characterize the particular gas under consideration. When a and b are zero, this equation becomes the perfect-gas law. It also approaches the perfect-gas law as V becomes larger.

Equation (10-4) is of the third degree in V; thus, for a given pressure and temperature, there may be three values of the volume which will satisfy the equation. Either all three of these may be real values, or one may be real and the other two complex. If two roots are complex, only the one real root has any physical significance. For a certain value of pressure and

temperature all three roots coincide. This point is called the *critical point.* For higher values of the pressure and temperature, there is only one real root.

Problem 12. Find the work done in compressing a gas which obeys the van der Waals equation of state from a volume V_1 to the volume V_2.

Problem 13. Take the temperature and the pressure as the independent variables of state, and draw the lines of constant volume for a gas which obeys the van der Waals equation.

Problem 14. Take the pressure and the volume as independent variables, and sketch the lines of constant temperature. Those portions of the curves which show three values of the volume for each value of temperature and pressure represent the condensation of the gas into the liquid. Show on this diagram the process necessary for transforming the gas into the liquid without going through the ordinary process of condensation.

Problem 15. Find the values of the critical pressure, temperature, and volume in terms of the constants a and b.

2. The First Law of Thermodynamics. *a. Statement of the Law.*—The first law of thermodynamics expresses the equivalence of heat and work as means of transferring energy. It states that the internal energy of a thermodynamic system can be increased in two ways. Either mechanical work can be done on the system, or energy may flow into it as heat from a body at a higher temperature.

The equation which states this law is

$$dU = \delta Q + \delta W \qquad \text{(I)}$$

where U represents the energy of the system. dU is the infinitesimal increase in energy of the system when an amount of energy δQ flows into it as heat and an amount of work δW is done on the system. The sign of δW is taken as positive when the surroundings do work on the system. It is often taken in the opposite way, since the conventions have been developed in the study of engines where the work done on the surroundings is the important thing. When the properties of the system form the subject of the investigation it seems a little better to take the sign as is done in (I).

The quantity U is a function of the state of the system only, and dU is therefore a differential of this quantity. dU can be integrated, and its integral between any two states is the difference between the energy of one state and the energy of the other. The path by which the system is taken from one state to the other is of no importance in evaluating this integral.

On the other hand, the quantities Q and W have no such significance. In equation (I), δQ and δW are written to indicate that only infinitesimal amounts of heat flow and of work are involved, but the sign δ does not imply a derivative. The integral of δQ or of δW can be evaluated only when the equation of state is known and when the path by which the system goes from one state to another is specified. The distinction between these two kinds of quantities is fundamental for the significance of the first law. The internal energy of a system can be increased in these two separate ways, and it is impossible to tell from the change of state of a system which way has been used.

The use of δQ added to δW implies that heat flow and work must be measured in the same units. The factor of proportionality between the thermometric unit of heat and the mechanical unit of heat is called the *mechanical equivalent of heat* J. It was the subject of very extensive experimental investigations by Joule and others. The best value at the present time is $J = 4.185$ joules per calorie. In this chapter heat flow will always be measured in joules unless otherwise specified.

Problem 16. Show that, if T and V are the independent variables which describe the state of the system and if the only work done by the surroundings is by the change in volume under the external pressure, the amount of energy which flows into the system in the form of heat during an infinitesimal change in state is

$$\delta Q = \left(\frac{\partial U}{\partial T}\right)_V dT + \left[\left(\frac{\partial U}{\partial V}\right)_T + p\right] dV \qquad (10\text{-}5a)$$

Problem 17. Show that, if the pressure and the volume are the independent variables,

$$\delta Q = \left(\frac{\partial U}{\partial p}\right)_V dp + \left[\left(\frac{\partial U}{\partial V}\right)_p + p\right] dV \qquad (10\text{-}5b)$$

Problem 18. Show that, if the pressure and the temperature are taken as the independent variables,

$$\delta Q = \left[\left(\frac{\partial U}{\partial p}\right)_T + p\left(\frac{\partial V}{\partial p}\right)_T\right] dp + \left[\left(\frac{\partial U}{\partial T}\right)_p + p\left(\frac{\partial V}{\partial T}\right)_p\right] dT \qquad (10\text{-}5c)$$

b. Applications of the First Law to Perfect Gases.—The equation of state given above for a perfect gas gives the relation between the pressure, the volume, and the temperature. It is necessary also to determine experimentally the dependence of the internal energy U upon the variables which define the state. Since the kinetic-theory model of a perfect monatomic gas gives for the energy

$$U = \tfrac{3}{2}nRT \text{ ergs} \qquad (10\text{-}6)$$

it may be taken as an additional part of the definition of an ideally perfect gas that $U = C_V T$, where C_V is a constant.

An early experiment performed by Joule showed that the ordinary gases approximate closely to this condition, in much the same way that they approximate closely to the equation of state (10-2). Joule filled a vessel with gas under a moderately high pressure. This vessel was connected to an evacuated vessel by means of a stopcock, and the whole was immersed in a calorimeter. When the stopcock was opened, the gas expanded without doing any work. It was also observed that the temperature of the water in the calorimeter did not change, so that no heat was absorbed in the process. The conclusion is that the internal energy of the gas is independent of the volume and is a function of the temperature only. Later and more accurate experiments by Joule and Thomson have shown that this result is approximate only, but it may be taken as a property of a perfect gas.

In dealing with the heat capacities of gases, it is important to differentiate between the heat capacity at constant volume C_V and the heat capacity at constant pressure C_p. The distinction must also be made for solids and liquids, but it is more important in gases because of their larger rates of thermal

expansion. The heat capacity at constant volume is the amount of energy that must be added to the substance to increase its temperature by one degree when it is not allowed to expand. Similarly, the heat capacity at constant pressure is the amount that must be added when the gas is allowed to expand just enough to maintain a constant pressure.

A distinction must also be made between heat capacity and specific heat. Heat capacity refers to any given quantity of substance and will be indicated by a capital C. Specific heat, on the other hand, refers to a specified quantity, such as a gram, and will be designated by a small letter c.

Problem 19. Show that

$$C_V = \left(\frac{\partial U}{\partial T}\right)_V \tag{10-7a}$$

Problem 20. Show that, for a perfect gas,

$$C_p = \left(\frac{\partial U}{\partial T}\right)_p + nR \tag{10-7b}$$

Problem 21. Show that

$$C_p = C_V + nR \tag{10-7c}$$

for a perfect gas.

Problem 22. Show that for a perfect gas, when the temperature and the volume are taken as the independent variables,

$$\delta Q = C_V\, dT + \frac{nRT}{V}\, dV \tag{10-8a}$$

Problem 23. Show that for a perfect gas, when the temperature and the pressure are the independent variables,

$$\delta Q = C_p\, dT - \frac{nRT}{p}\, dp \tag{10-8b}$$

Problem 24. Show that for a perfect gas, when the pressure and the volume are the independent variables,

$$\delta Q = \frac{C_V}{nR}\, V\, dp + \frac{C_p}{nR}\, p\, dV \tag{10-8c}$$

If the path by which a change of the state of the system takes place is given, equations (10-8) can be integrated. There

are several types of change which it is convenient to use. One is that in which one of the independent variables is held fixed. In this way paths of constant pressure, volume, or temperature are obtained. Another type of path is the adiabatic. When a system is so arranged that it can absorb or give out no energy in the form of heat, any changes that take place are said to be *adiabatic.* The equation of these paths can be obtained by setting $\delta Q = 0$.

Consider the case of an adiabatic transformation of a perfect gas whose state is described by pressure and volume. Equation (10-5*b*) gives the relationship between dp and dV for an adiabatic change

$$\frac{dp}{dV} = -\frac{(\partial U/\partial V)_p + p}{(\partial U/\partial p)_V}$$

From the definition of a perfect gas

$$U = C_V T = \frac{C_V pV}{nR}$$

Hence

$$\left(\frac{\partial U}{\partial V}\right)_p = \frac{C_V}{nR} p \qquad \left(\frac{\partial U}{\partial p}\right)_V = \frac{C_V}{nR} V$$

From this and equation (10-7*c*)

$$\frac{dp}{dV} = -\frac{(C_V + nR)p}{C_V V} = -\frac{C_p}{C_V}\frac{p}{V}$$

and by integrating this equation it can be shown that

$$pV^\gamma = \text{const} \tag{10-9a}$$

where $\gamma = C_p/C_V$.

Problem 25. Show that for an adiabatic expansion of a perfect gas

$$TV^{\gamma-1} = \text{const} \tag{10-9b}$$

and

$$Tp^{(1-\gamma)/\gamma} = \text{const} \tag{10-9c}$$

To show the way in which the internal energy U can be a function of the state of a system at the same time that the work done on it and the heat flow into it depend on the processes carried out, it is instructive to consider a special case. Consider

a perfect gas at the temperature T' with the volume V'. Let it first be warmed to the temperature T'' at constant volume. The heat flow into the gas is then, from equation (10-5a)

$$Q = \int_{T'}^{T''} \left(\frac{\partial U}{\partial T}\right)_V dT = C_V(T'' - T') \qquad (10\text{-}10a)$$

Then, at the constant temperature T'', let the gas expand from V' to V''. Again from equation (10-5a) the heat flow into the gas is

$$Q = \int_{V'}^{V''} \left[\left(\frac{\partial U}{\partial V}\right)_T T + p\right] dV = nRT'' \int_{V'}^{V''} \frac{dV}{V}$$
$$= nRT'' \log \frac{V''}{V'} \qquad (10\text{-}10b)$$

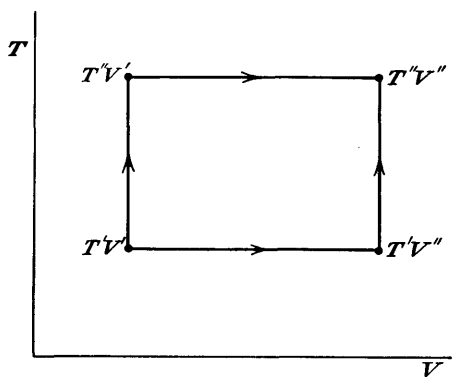


FIG. 10-1.—A cycle carried out at constant temperature and constant volumes.

For a perfect gas, $(\partial U/\partial V)_T = 0$; hence the only contribution to the integral comes from $\int p\, dV$.

If the change in state is made along the other path shown in Fig. 10-1, the heat flow during the first leg is

$$Q = nRT' \log \frac{V''}{V'} \qquad (10\text{-}10c)$$

and during the second leg is

$$Q = C_V(T'' - T') \qquad (10\text{-}10d)$$

This example shows that the heat flow depends on the path in the T-V plane and is not at all the same for all paths.

It is also of interest to compute the work done on the gas

in making this change of state. The only work done is associated with the change in volume and is $-\int p\,dV$.

$$W = -\int_{V'}^{V''} nRT'' \frac{dV}{V} = -nRT'' \log \frac{V''}{V'} \quad (10\text{-}11a)$$

for the first path and

$$W = -nRT' \log \frac{V''}{V'} \quad (10\text{-}11b)$$

for the second path.

From the first law of thermodynamics it follows that

$$U(T'',V'') - U(T',V') = Q + W = C_V(T'' - T') \quad (10\text{-}12)$$

If it is assumed known that C_V is independent of V so that the heat capacity is the same no matter whether the heating is done with the volume V' or V'', this equation serves as an illustration of the first law. On the other hand, the first law can be used in this way to show that the specific heat is independent of the volume.

Problem 26. Consider a perfect gas which undergoes a cyclic process made up of elements at constant temperature and at constant volume. Let it first expand at the temperature T' from the volume V' to the volume V''. Then let it be heated at the volume V'' to the temperature T'', compressed at the temperature T'' to the volume V', and finally cooled at V' to T'. Compute the heat absorbed and the net work done on the gas.

Problem 27. Compute the work done on a gas that obeys the van der Waals equation when it is compressed from V_1 to V_2 at constant temperature.

Problem 28. Let the state of a perfect gas be represented by the independent variables p and V. Compute the heat absorbed and the work done on the gas when it is changed from the state $(p'V')$ to the state $(p''V'')$. It is convenient to make the path out of one section at constant pressure and another at constant volume.

Cyclical processes similar to the one in Prob. 26 are frequently used in thermodynamic reasoning. One of particular usefulness is illustrated in Fig. 10-2. The substance may be considered to start at T_1V_1 and to expand at constant temperature to T_1V_2. To maintain the temperature constant during this expansion, heat must be supplied to the working

substance. In the case of a perfect gas this amount of heat is

$$(T_1V_1 \rightarrow T_1V_2) \qquad Q = \int_{V_1}^{V_2} \left[\left(\frac{\partial U}{\partial V} \right)_T + p \right] dV$$
$$= nRT_1 \log \frac{V_2}{V_1} \quad (10\text{-}13a)$$

The quantity $(\partial U/\partial V)_T = 0$ for a perfect gas. The work done on the gas during this process is

$$(T_1V_1 \rightarrow T_1V_2) \qquad W = -\int_{V_1}^{V_2} p\, dV = -nRT_1 \log \frac{V_2}{V_1} \quad (10\text{-}13b)$$

The sum of these two quantities is zero as it must be, since the

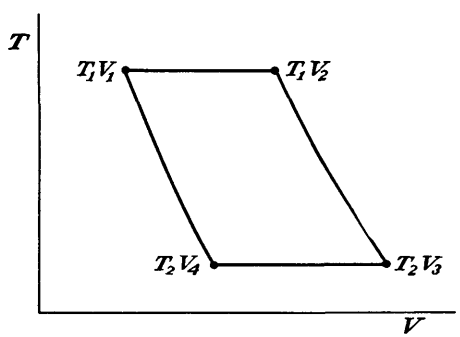


FIG. 10-2.—A cycle carried out isothermally and abiabatically.

internal energy of a perfect gas depends only on the temperature and is the same at (T_1V_1) as at (T_1V_2).

The next step in the process is an adiabatic expansion from T_1V_2 to T_2V_3. Along this line $\delta Q = 0$ by the definition of an adiabatic process, and by equation (10-9b) $TV^{\gamma-1} = K$. Hence the work can be calculated.

$$(T_1V_2 \rightarrow T_2V_3) \qquad W = -\int p\, dV = -nR \int_{V_2}^{V_3} T \frac{dV}{V}$$
$$= -nRK \int_{V_2}^{V_3} \frac{dV}{V^\gamma}$$
$$= \frac{nRK}{\gamma - 1} \left(\frac{1}{V_3^{\gamma-1}} - \frac{1}{V_2^{\gamma-1}} \right)$$
$$= \frac{nR}{\gamma - 1} (T_2 - T_1) \quad (10\text{-}13c)$$

The third step is an isothermal compression from T_2V_3 to T_2V_4. The heat and work are

$$(T_2V_3 \rightarrow T_2V_4) \qquad Q = nRT_2 \log \frac{V_4}{V_3} \tag{10-14a}$$

$$(T_2V_3 \rightarrow T_2V_4) \qquad W = -nRT_2 \log \frac{V_4}{V_3} \tag{10-14b}$$

These two quantities are equal and opposite, so that the internal energy U does not change during this process.

The fourth step is an adiabatic compression during which no heat flows and

$$(T_2V_4 \rightarrow T_1V_1) \qquad W = \frac{nR}{\gamma - 1}(T_1 - T_2) \tag{10-14c}$$

This step returns the gas to its original state and its original internal energy. The sum of the heat absorbed and the work done on the gas must be zero. In this case the net heat absorbed was

$$Q = nR\,(T_1 - T_2) \log \frac{V_2}{V_1} \tag{10-15}$$

since

$$\frac{V_2}{V_1} = \frac{V_3}{V_4}$$

The total work done on the gas is the negative of this. The gas does work on the surroundings, and the energy for this work gets into the gas by heat flow.

Problem 29. Compute the work done along an adiabatic path in the T-p plane in changing the state of a perfect gas from T_1p_1 to T_2p_2.

3. The Second Law of Thermodynamics. *a. Statement of the Second Law.*—The first law of thermodynamics states the equivalence of heat and work as methods of transferring energy; the second law states the properties that distinguish the transfer of energy by heat flow from transfer by work. The second law may be stated as follows: It is impossible for a self-acting machine, unaided by any external agency, to convey heat from one body to another at a higher temperature. This is the form in which it was stated by Clausius. The expressions used show clearly the persistence of the idea of heat as an

indestructible substance. Possibly a modification of this statement which would better accord with the mechanical theory of heat is: *It is impossible for any self-acting machine, unaided by any external agency, to absorb energy in the form of heat from one body and to give up energy in the form of heat to another body at a higher temperature.*

A great many different formulations of this second law have been given. Attempts have also been made to prove it from the first law or from some proposition thought to be "self-evident." It is certainly impossible to prove it from the first law because it states facts which are not included in the first law. Whether the statements with which it is equivalent are self-evident or not depends entirely upon the state of mind of the person stating them. The above statement of the second law is made as a generalization from observation, and its truth or falsity can be determined only by experiment. Thus far it has proved to be correct in the macroscopic sense in which it is meant. The microscopic interpretation of this law in terms of statistical mechanics will be reserved until the next chapter.

On the basis of the law as just stated, it is possible to show that *all reversible engines have the same efficiency and that no engine can have an efficiency higher than that of a reversible engine.* A heat engine, or an engine, is a device that absorbs energy in the form of heat and does work on its surroundings. If the engine is 100 per cent efficient, all the energy absorbed is transformed into work done on the surroundings. When the efficiency is less than 100 per cent, some of the energy absorbed as heat is given out again as heat at a lower temperature. Since the energy for doing the work is to be obtained only by absorption as heat, the engine must work in cycles. At the end of a cycle, all parts of the engine will be in exactly the same state as at the beginning. Some energy will have been absorbed at one temperature, some work will have been done on the surroundings, and the remainder of the absorbed energy will have been given out as heat at a lower temperature. The engine is only the agent for effecting the transformation of energy and suffers no change itself.

This reversibility is a very important thermodynamic con-

cept. In this case the implication is that the engine can be driven backward by doing on it the work it normally does on its surroundings. The working substance then goes through its cycle in the opposite direction; it absorbs some heat at the lower temperature and gives out a larger quantity of heat at the higher temperature. As applied to other processes, the term represents the fact that the conditions describing the process do not specify the direction in which it will go. If a gas is expanding reversibly at a pressure p, it is implied that the pressure against which it is working is also p, and an infinitesimal increase in the external pressure would turn the expansion into a compression. In either case there is no net force to produce an acceleration. In the case of reversible heat flow, both bodies must be at essentially the same temperature so that the flow can be directed one way or the other by an infinitesimal change in temperature.

The cycles described and analyzed above in discussing the first law are typical of reversible cycles that can be carried out with perfect gases. The efficiency of such a cycle is defined as the ratio of the work done on the surroundings to the heat absorbed at the higher temperature. This ratio, for reversible cycles, is independent of the nature of the working substance and the nature of the cycle and is a function of the working temperatures only. Hence it can be evaluated by using a particular cycle whose properties are known.

Problem 30. Show from the second law of thermodynamics that all reversible heat engines have the same efficiency and that no engine can have a greater efficiency than a reversible engine. This can be shown by allowing the more efficient engine to drive the other backwards, which would make possible a violation of the second law.

Problem 31. From the result of Prob. 29 show that the efficiency of a reversible engine which absorbs energy as heat at the temperature T_1 and gives out energy as heat at the temperature T_2 is $(T_1 - T_2)/T_1$.

Problem 32. Show that, for any reversible cycle in which an amount of heat Q_1 is absorbed at the temperature T_1 and $-Q_2$ is given out at the temperature T_2,

$$\frac{Q_1}{T_1} + \frac{Q_2}{T_2} = 0 \tag{10-16}$$

The heat absorbed is given the positive sign, while the heat given out is considered as negative.

Problem 33. Show that for any reversible cycle

$$\int \frac{\delta Q}{T} = 0 \tag{10-17}$$

where the integration is around the cycle.

b. The Absolute Scale of Thermodynamic Temperature.—The second law of thermodynamics provides a method of defining temperature which is independent of the properties of any substance. It was suggested by Lord Kelvin that, since the efficiency of a reversible cycle is a function of the temperature only, it could be used to define temperature. Such a definition establishes a scale with which all actual thermometers may be calibrated.

The characteristic of all reversible cycles is that they absorb energy at one temperature T_1 and give out energy in the form of heat at another temperature T_2. They also deliver an amount of work which is equal to the difference between the energy absorbed and the energy given out as heat. Since the quantities of energy can be measured, it is possible to define the temperatures between which the cycle works in such a way that

$$\frac{Q_1}{T_1} + \frac{Q_2}{T_2} = 0$$

This will be independent of the nature of the cycle because all cycles will have the same efficiency.

In the previous work the temperature has been implicitly defined on the basis of the perfect-gas law. Since, however, there exists no perfect gas, it is impossible to prescribe a method for establishing such a scale. The above given definition of temperature is in complete accord with the definition on the basis of a perfect gas but is independent of the existence of such a gas. When temperature is used in thermodynamics, it is always referred to this absolute, or Kelvin, scale.

c. Entropy.—The results of Probs. 32 and 33 make it possible to define a function S called the *entropy*. The entropy is a

function of the state of the system and is given when the independent variables of state, such as pressure and volume, are specified. The entropy is defined by defining its differential along any infinitesimal portion of path. For a reversible process this is

$$dS = \frac{\delta Q}{T} \tag{10-18}$$

The quantity on the left side of this equation is an exact differential because the quantity on the right has been shown in Prob. 33 to be an exact differential when the change of state is carried out reversibly. It is because the integral of dS around any cycle is zero that S is a single-valued function of the state.

The absolute value of the entropy is not defined by equation (10-18). If some state is taken as a standard, it is possible to determine the entropy of any other state by integrating equation (10-18) along a curve which represents a reversible passage from the standard state to the state in which it is desired to know the entropy.

If a system passes from one state to another by a process which is not reversible, the change in entropy can be calculated from a knowledge of the possible reversible processes, but it will not be equal to $\delta Q/T$ for the irreversible process. This may be seen by considering the system to pass from the state S_1 to the state S_2. The change in entropy is defined, and the change in internal energy is defined, since these two are functions of the state of the system only. When the process is reversible, the amount of work done by the system can be obtained from the first law in terms of the heat absorbed and this is the maximum amount of work. If the process is irreversible, the amount of work is less than the maximum and, according to the first law, the heat absorbed is also less. Hence, in an irreversible isothermal process the change in entropy is greater than the integral of $\delta Q/T$.

A mathematical formulation of the second law of thermodynamics is that for all processes

$$dS \geq \frac{\delta Q}{T} \tag{II}$$

If, then, a system is thermally insulated so that no heat can enter or leave it, the entropy can only increase or remain constant.

To compute the entropy of a perfect gas, consider n moles of the gas to be at the pressure p_0, temperature T_0, and volume V_0. First heat it to the temperature T while holding the volume constant. The pressure will increase; but since the volume is constant, no work is done and the internal energy increases directly with the temperature. The heat absorbed in each element of this process is $\delta Q = C_V\,\delta T$ so that

$$S(V_0,T) - S(V_0,T_0) = \int_{T_0}^{T} C_V \frac{dT}{T} = C_V \log \frac{T}{T_0}$$

If then the temperature is held constant and the gas is permitted to expand from V_0 to V at constant temperature the heat absorbed is just equal to the work done by the gas so that

$$S(V,T) - S(V_0,T) = \frac{1}{T}\int_{V_0}^{V} p\,dV = \frac{nRT}{T}\int_{V_0}^{V}\frac{dV}{V} = nR \log \frac{V}{V_0}$$

Combining these two results gives the entropy as a function of V and T and an arbitrary value assigned to the entropy at V_0T_0.

$$S(V,T) = nR \log \frac{V}{V_0} + C_V \log \frac{T}{T_0} + S(V_0,T_0) \quad (10\text{-}19)$$

This computation of entropy involves a knowledge of the equation of state and of the internal energy U of the substance as a function of the variables describing the state.

Problem 34. Show that when a perfect gas is allowed to expand into a vacuum the increase in entropy is greater than would be computed on the basis of the heat absorbed.

Problem 35. Two bodies at different temperatures are put in contact. Assume that each body has the same temperature throughout, and show that the entropy increases when δQ flows from one to the other.

Problem 36. If the internal energy of a gas that obeys the van der Waals equation of state is given by

$$U = C_V T - \frac{n^2 a}{V} + w$$

where w is a constant, compute its entropy as a function of T and V.

4. Application of the Two Laws. *a. Various Functions of State.*—Thus far the energy and the entropy have been defined as functions of the state of the system, together with the variables p, V, and T. Any two of these five quantities may be taken as the independent variables to specify the state, and the others may be expressed in terms of them. There must then be three "equations of state" to give the three dependent quantities in terms of the two independent variables, and these equations must be determined experimentally for any particular substance. The laws of thermodynamics, however, give some relationships between them.

The specification of the state of a system by only two variables is sufficient for simple systems. In some cases more quantities are required, but the principles can be described in terms of the simplest cases.

If the only way in which the surroundings can do work upon the system is through a uniform external pressure, the first and second laws of thermodynamics can be combined into an equation involving only the variables whose values are given for a fixed state. By means of equation (10-18), the first law can be written

$$dU = T\,dS - p\,dV \tag{10-20}$$

Because of the use of equation (10-18) this holds for reversible processes only. If S and V are taken as the independent variables, the condition that U is a function of state and dU is an exact differential leads to

$$\left(\frac{\partial T}{\partial V}\right)_S = -\left(\frac{\partial p}{\partial S}\right)_V \tag{10-21}$$

This process of deriving relations between differential quotients of the variables of state from the condition for an exact differential is characteristic of much thermodynamic reasoning.

By the use of the combination of the first and second laws, it is possible to determine $(\partial U/\partial V)_T$ from the equation that gives p as a function of V and T. This gives some information about U as a function of V and T, although it does not determine the function entirely. If T and V are taken as the independent

variables, equation (10-20) may be written

$$\left(\frac{\partial U}{\partial T}\right)_V dT + \left(\frac{\partial U}{\partial V}\right)_T dV = T\left(\frac{\partial S}{\partial T}\right)_V dT + T\left(\frac{\partial S}{\partial V}\right)_T dV - p\,dV \quad (10\text{-}22)$$

By equating coefficients of the independent differentials dT and dV it follows that

$$\left(\frac{\partial U}{\partial T}\right)_V = T\left(\frac{\partial S}{\partial T}\right)_V \quad (10\text{-}22a)$$

and

$$\left(\frac{\partial U}{\partial V}\right)_T = T\left(\frac{\partial S}{\partial V}\right)_T - p \quad (10\text{-}22b)$$

Differentiating equation (10-22a) with respect to V leads to

$$\frac{\partial^2 U}{\partial V\,\partial T} = T\frac{\partial^2 S}{\partial V\,\partial T} \quad (10\text{-}22c)$$

where the subscript indicating the independent variable held constant is omitted, since V and T are the two independent variables and the order of differentiation does not matter. Similarly, differentiating (10-22b) with respect to T leads to

$$\frac{\partial^2 U}{\partial V\,\partial T} = T\frac{\partial^2 S}{\partial V\,\partial T} + \left(\frac{\partial S}{\partial V}\right)_T - \left(\frac{\partial p}{\partial T}\right)_V \quad (10\text{-}22d)$$

Comparing equations (10-22c) and (10-22d) shows that

$$\left(\frac{\partial S}{\partial V}\right)_T = \left(\frac{\partial p}{\partial T}\right)_V \quad (10\text{-}22e)$$

so that equation (10-22b) becomes

$$\left(\frac{\partial U}{\partial V}\right)_T = T\left(\frac{\partial p}{\partial T}\right)_V - p \quad (10\text{-}23)$$

The quantities on the right side of this equation can be evaluated from the equation of state that gives p as a function of T and V so that $(\partial U/\partial V)_T$ for any volume and temperature can be determined.

Problem 37. Show from equation (10-23) that the energy of a perfect gas is independent of the volume.

Problem 38. Find the volume dependence of the internal energy of a gas that follows the van der Waals equation.

Problem 39. Show by differentiating equation (10-23) that

$$\left(\frac{\partial C_V}{\partial V}\right)_T = T\left(\frac{\partial^2 p}{\partial T^2}\right)_V \tag{10-24}$$

and show that, for both a perfect gas and a van der Waals gas, the specific heat at constant volume is independent of the temperature.

There are a number of other functions of the state of a system that are frequently useful.

$$F = U - TS \tag{10-25}$$

is often called the *free energy* but is perhaps more appropriately called the *work function.*

$$H = U + pV \tag{10-26}$$

was originally called the *heat content* or *heat function.* The significance of the name was based on the old idea of heat as an indestructible fluid. The modern name *enthalpy* is more appropriate.

$$G = U - TS + pV \tag{10-27}$$

is called the *thermodynamic potential* at constant pressure or the *Gibbs function.* All three of these functions are useful because they are functions of state and their differentials are exact.

Problem 40. Show that

$$dF = -S\,dT - p\,dV \tag{10-28}$$

Problem 41. Show that

$$dH = T\,dS + V\,dp \tag{10-29}$$

Problem 42. Show that

$$dG = -S\,dT + V\,dp \tag{10-30}$$

Problem 43. Show that

$$S = -\left(\frac{\partial F}{\partial T}\right)_V \quad \text{and} \quad p = -\left(\frac{\partial F}{\partial V}\right)_T \tag{10-31}$$

Problem 44. Show that

$$\left(\frac{\partial S}{\partial V}\right)_T = \left(\frac{\partial p}{\partial T}\right)_V \tag{10-32}$$

Problem 45. Show from the above equation that

$$L = T\left(\frac{\partial p}{\partial T}\right)_V (V_2 - V_1) \tag{10-33}$$

where L is the heat of change of state and V_1 is the volume of one mole of the substance in the initial state while V_2 is the volume in the final state. T is the temperature at which the change of state takes place.

Problem 46. Show that

$$\left(\frac{\partial T}{\partial p}\right)_V = \frac{T(V_2 - V_1)}{L} \tag{10-34}$$

and calculate the amount of the depression of the melting point of ice by a pressure of 1 atmosphere.

b. Work Done in a Reversible Cycle.—It is frequently desirable to use graphical methods of representing the states of a system and various changes in state. If the pressure and the volume are used as the coordinates in which a cyclical process is described, the area inside the curve which represents the path of the system gives the work done on it. If other variables are used, the work is not directly given by the area but is equal to the integral of a function γ multiplied by an element of area dA and integrated over the area inside the curve. This can be shown by drawing isothermal and adiabatic curves close together. The work done in any small cycle bounded by two adiabatics and two isothermals is proportional to the area. Hence the whole work can be given by the above integral.

Problem 47. Show that the work done in any cyclical process, when x and y are the independent variables of state, is given by

$$W = \int \gamma \, dA$$

where

$$\gamma = \frac{\partial p}{\partial x}\frac{\partial V}{\partial y} - \frac{\partial p}{\partial y}\frac{\partial V}{\partial x} \tag{10-35}$$

Problem 48. Use temperature and entropy to represent graphi-

cally a cycle which consists of adiabatics and isothermals. Compute the work done and the heat absorbed.

Problem 49. Use pressure and volume as the independent variables to represent graphically a cycle which consists of the evaporation of a liquid at constant temperature, an adiabatic expansion and evaporation, an isothermal condensation, and an adiabatic compression. Compute the work done and the heat absorbed.

Problem 50. Use entropy and temperature as independent variables in the above problem.

c. The Joule-Thomson Porous-plug Experiment.—In 1845 Joule showed that the energy of a gas is almost independent of its volume by permitting a fixed mass of gas to expand into a vacuum and observing that there was no change in temperature. The conditions of the experiment were such that the accuracy was not very great, and therefore Joule and Thomson later revised the experiment to permit more precision. The gas was allowed to flow through a tube containing a plug of porous material. This plug retarded the flow of gas and maintained the pressure on the input side higher than on the outflow side. The temperature on the two sides was measured by means of sensitive thermometers, and it was established that a small temperature difference did exist.

The whole process was carried out adiabatically, so that the only change in internal energy was due to mechanical work done on the gas. As a given mass of gas is pushed into the plug at the pressure p_1, the work done on it is p_1V_1, where V_1 is the volume at the pressure p_1 and the temperature of the experiment. As the gas comes out of the plug against the pressure p_2, it will expand to the volume V_2 and do work against the surroundings of amount p_2V_2. The internal energy after the expansion will then be

$$U_2 = U_1 + p_1V_1 - p_2V_2$$

or

$$H = U + pV = \text{const} \tag{10-36}$$

The experiment is thus carried out at constant enthalpy.

With pressure and temperature as the independent variables of state and the specification that the enthalpy is constant,

$$dH = 0 = T\left(\frac{\partial S}{\partial T}\right)_p dT + T\left(\frac{\partial S}{\partial p}\right)_T dp + V\,dp \quad (10\text{-}37)$$

and

$$\left(\frac{\partial T}{\partial p}\right)_H = -\frac{T(\partial S/\partial p)_T + V}{T(\partial S/\partial T)_p} \quad (10\text{-}38)$$

From equation (10-38) it is possible to compute the temperature change to be expected in the porous-plug experiment when the entropy is known as a function of the pressure and the temperature.

Problem 51. Show that a perfect gas experiences no change in temperature in the porous-plug experiment.

Problem 52. Work out the temperature change in the porous-plug experiment when the gas follows the van der Waals equation of state and its internal energy is as given in Prob. 36.

Problem 53. Show that

$$T\left(\frac{\partial S}{\partial T}\right)_p = C_p \quad (10\text{-}39)$$

Equation (10-39) shows that equation (10-38) is equivalent to

$$C_p\left(\frac{\partial T}{\partial p}\right)_x = -T\left(\frac{\partial S}{\partial p}\right)_T - V$$

It also follows from equation (10-30) and the fact that dG is an exact differential that

$$\left(\frac{\partial S}{\partial p}\right)_T = -\left(\frac{\partial V}{\partial T}\right)_P$$

As a result

$$T = \frac{V + C_p(\partial T/\partial p)_H}{(\partial V/\partial T)_H} \quad (10\text{-}40)$$

It was pointed out by Lord Kelvin that this equation permits the calibration of actual thermometers in terms of the absolute thermodynamic temperature T.

CHAPTER XI

STATISTICAL MECHANICS

According to the present view of the constitution of matter, all bodies are made up of large numbers of similar particles called *molecules*. In different cases these may be chemical atoms or molecules, or they may be electrons and atomic nuclei. Individually they can be treated to a large extent as small solid bodies, which obey the laws of mechanics. The properties of the larger bodies are then to be explained as consequences of the motions and the interactions of the molecules. This is, for example, the idea of the mechanical theory of heat. It suggests the possibility of deducing the laws of thermodynamics from the laws of mechanics applied to a large number of particles. The study of statistical mechanics is concerned with a formulation of the properties of large aggregations of particles in terms of their individual properties and the attempt to explain the properties of matter in this way.

In this chapter, only the classical statistical mechanics will be treated; *i.e.*, the properties attributed to the molecules will be those of ordinary matter as it is known directly to our senses. These are not, however, the true properties of the molecules, and the modifications that must be made in the treatment lead to the quantum statistical mechanics. In many cases, however, the differences between classical and quantum statistical mechanics are in detail rather than in the general principles.

1. The Phase Space.—Problems in statistical mechanics are nearly always treated by means of Hamilton's equations of motion. The solutions of these equations give the values of the coordinates q_i and their conjugate momenta p_i as functions of the time and the initial values. If, for a system of one coordinate, the value of the coordinate is plotted along a horizontal axis and the value of the conjugate momentum is plotted along a

vertical axis, a point in this p-q plane will represent a state of the system. As the values of p and q change with the time, the representative point will describe an orbit in the plane.

The space in which the representative point moves is called the *phase space.* For a system with one degree of freedom, or one independent coordinate, it is a phase plane. For a system

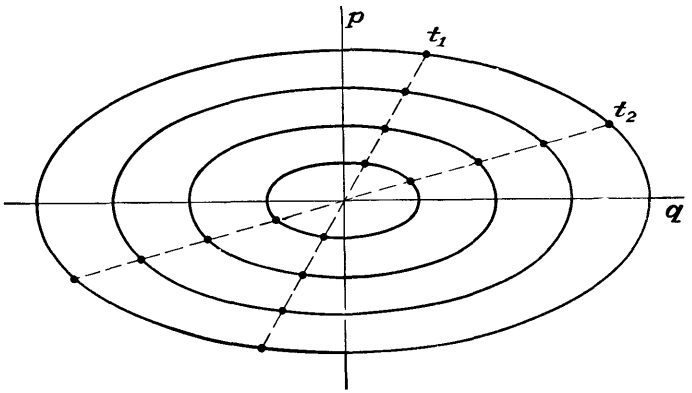


FIG. 11-1.—Orbits in the phase space of the point representing a simple oscillator.

with N coordinates the phase space has $2N$ dimensions, and the position of the single representative point in this $2N$-dimensional phase space gives all the possible information about the system. The point represents its state or phase.

Only for one-dimensional systems can the phase space be easily represented graphically. In the case of a one-dimensional-harmonic oscillator

$$\begin{aligned} q &= A \sin (\omega t - \epsilon) \\ m\dot{q} &= p = m\omega A \cos (\omega t - \epsilon) \end{aligned} \tag{11-1}$$

The orbit described by the representative point in this phase plane is given by the relationship between q and p.

$$\frac{q^2}{A^2} + \frac{p^2}{m^2\omega^2 A^2} = 1 \tag{11-2}$$

This is an ellipse whose semiaxes are A and $m\omega A$. These axes are determined by the initial values of q and p for

$$A^2 = q^2 + \frac{p^2}{m^2\omega^2} = q_0^2 + \frac{p_0^2}{m^2\omega^2} \tag{11-3}$$

The value of the constant ϵ does not affect the size or shape of the orbit, but it does affect the position in the orbit as a function of the time.

If at the time $t = t_0$ a number of different identical oscillators have $q = 0$ but values of $\dot{q}$ lying between v and $-v$, the representative points will lie on a straight line coinciding with the p axis. As time goes on, these points will describe their individual orbits but will continue to lie on a straight line.

It is also interesting to note that no two of these orbits intersect. A point on one never gets on another. Furthermore, each of these orbits can be completely characterized by the total energy corresponding to it.

Problem 1. Determine the orbit in the phase plane of the point that represents a particle projected vertically upward against gravity. Find the lines of constant total energy in this plane, and show that the point moves along one of them. Indicate for several different times the position of a number of particles projected upward at the same time with different velocities.

2. Distribution in Phase and Liouville's Theorem.—As has been mentioned before, the solution of the differential equations of mechanics gives the coordinates and the momenta of the system in terms of the initial conditions and the time. It is necessary to know these initial conditions before the configuration of the system can be predicted. In physical systems in which the number of degrees of freedom, or the number of independent coordinates, is very large, it is impossible to determine these initial conditions experimentally and it would be impossible to use the information even if it could be obtained. In the case of a gas, the knowledge of the initial conditions would involve the knowledge of the position and the momentum of every molecule at some one time. It is obviously impractical to get this information, since only relatively coarse means of studying the gas are available, and, because of the large number of molecules, it would be impossible to carry out the necessary computations in any reasonable time. It is necessary, then, to see what can be learned from the molecular theory of matter without an exact knowledge of the initial conditions.

If the initial conditions of the system were known, it would be possible to place a point in the phase space at the proper location to represent these conditions. The future motion of the point would then be given by the differential equations of motion, and the future behavior of the system could be predicted. Since, however, it is not known where the point should be put, the next best thing is to put a point at every place that might be the correct place. This leads to the study of an *ensemble* of systems. Instead of studying a single system whose initial conditions are known, the statistical procedure is to study a whole collection of systems whose initial conditions are distributed in some way throughout the phase space. Each point in the phase space represents a separate system, and each point, therefore, carries out its motion independently of all the others. Instead of studying the motion of a single point, one studies the streaming of the whole ensemble of points.

The motions of the mechanical systems will be governed by the mutual forces between the component particles and also by the values of a number of outside parameters. These outside parameters, or coordinates, will have at any time the same values for all the systems of the ensemble, although they may change with the time. If the system is a gas contained in a cylinder closed by a piston, the position of the piston will be the same for all the gases of the ensemble. Thus a variation in the external parameters will cause a change of the whole ensemble.

It is convenient to consider the representative points to be so densely distributed throughout the phase space that it is possible to speak of a density in phase D. This implies, of course, that the volume elements by means of which the density is calculated will not be allowed really to approach zero as a limit but will always be kept so large as to contain a very large number of points. On the other hand, the elements must be made so small that the density does not change appreciably from one to the next. Then the number of systems whose representative points lie in the element of the phase space $dq_1 \cdots dq_n\, dp_1 \cdots dp_n$ will be (by the definition of D) $D\, dq_1 \cdots dq_n\, dp_1 \cdots dp_n$.

A rather remarkable theorem, called *Liouville's theorem* or the *theorem of the conservation of density in phase,* can be proved about this density D. Consider the surface element perpendicular to the axis of q_1 and on the side of the volume element toward the lower values of q_1. The number of systems that pass through this surface per unit time into the volume element under consideration will be

$$D\dot{q}_1\, dq_2 \cdots dq_n\, dp_1 \cdots dp_n$$

The number that pass out through the opposite surface will be

$$\left[D\dot{q}_1 + \frac{\partial(D\dot{q}_1)}{\partial q_1}\, dq_1\right] dq_2 \cdots dq_n\, dp_1 \cdots dp_n$$

Similar expressions will hold for all the other surfaces. By combining all these expressions, one obtains the result that the net rate of increase of the density of representative points within this element of volume of the phase space is given by

$$\frac{\partial D}{\partial t} = -\sum_i \frac{\partial}{\partial q_i}(D\dot{q}_i) - \sum_i \frac{\partial}{\partial p_i}(D\dot{p}_i) \qquad (11\text{-}4)$$

From the Hamiltonian equations of motion

$$\frac{\partial \dot{q}_i}{\partial q_i} = -\frac{\partial \dot{p}_i}{\partial p_i}$$

so that

$$\frac{dD}{dt} = \frac{\partial D}{\partial t} + \sum_i \frac{\partial D}{\partial q}\dot{q}_i + \sum_i \frac{\partial D}{\partial p}\dot{p}_i = 0 \qquad (11\text{-}5)$$

Equation (11-5) is called Liouville's theorem. It is the fundamental theorem of statistical mechanics.

A simple illustration of this theorem can be given in the case of a particle projected upward against gravity. Figure 11-2 represents the phase plane with the coordinate q plotted horizontally and the momentum plotted vertically. At $t = 0$ let four representative points be at the corners of the square, 1, 2, 3, 4. Points 1 and 2 have $q = 0$ with different values of p. Points 3 and 4 have the same small positive value of q, and the same two values of p as points 1 and 2. After a time these points occupy the positions 1′. 2′, 3′, 4′. The difference in momentum

between points 1′ and 4′ and points 2′ and 3′ is just the same as between points 1 and 4, since the momentum of all points decreases at the same rate. Furthermore, the difference in q between points 1′ and 4′ and between 2′ and 3′ is just the same

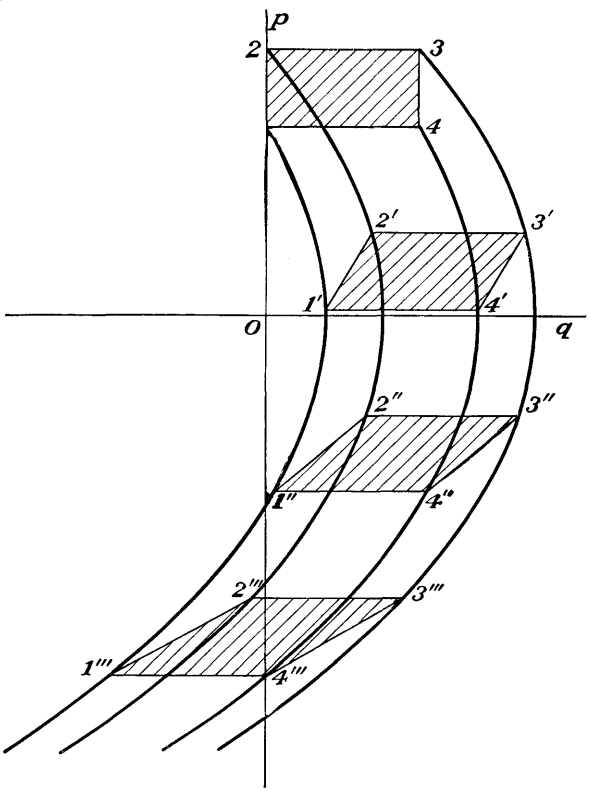


FIG. 11-2.—Orbits in the phase space of points representing a particle projected upward against gravity. The area enclosed by the four points remains constant as they move.

as between 1 and 4. The area of the parallelogram is just the same as that of the original square. As time goes on, the angles of the parallelogram change but the area remains constant.

Problem 2. Consider the phase plane for a particle moving in a straight line. Show that the area enclosed by the lines connecting a number of points does not change as time goes on.

3. Statistical Equilibrium and the Canonical Distribution.—A distribution of points in the phase space is said to be in *statistical equilibrium* if the partial derivative of the density

with respect to the time is zero. This means that the density of the representative points at each point in the phase space is constant. If the density were uniform, the distribution would be in statistical equilibrium, since all the derivatives would be zero.

There are other distributions, however, which are also in equilibrium. If the density in phase is a function of constants of the motion only, the distribution will be in equilibrium. Constants of the motion are such things as energy, momentum, angular momentum, and other quantities that, in any particular case, may remain constant as the system carries out its motion. Each constant of the motion defines a set of surfaces in the phase space on each member of which it has a specified value. The intersection of the surfaces for different constants defines the path of the point representing the system. A point will never leave the surface defined by one of the constants of its motion. Hence, if the distribution is such that it depends only upon these constants, it will be uniform over the surfaces representing this constant and will remain uniform over these surfaces. The whole will then be a distribution in statistical equilibrium. In a conservative system the total energy is a constant of the motion. If the density D is a function of the energy only $\partial D/\partial t = (\partial H/\partial t)(dD/dH) = 0$. This is one of the most useful kinds of equilibrium distributions.

Instead of dealing with the density in phase D, it is often more convenient to deal with the probability of phase P. P is defined as the fraction of the total number of points per unit volume of the phase space. Hence, $P = D/N$, where N is the total number of systems. Although both D and N may be thought of as infinite in order to get a continuous distribution, P will be finite. P must satisfy the condition

$$\int P\, dq_1\, dq_2 \cdots dq_n\, dp_1 \cdots dp_n = 1 \qquad (11\text{-}6)$$

where the integration is over the whole of the phase space, *i.e.*, over the whole range of each of the quantities q_i and p_i.

Since the p_i's have an infinite range, this condition puts certain limitations on the nature of the function P. It is not

possible to have P constant without having $P = 0$ at all points. If the distribution is to be one of equilibrium because the density is a function of the energy only, it is impossible to have P proportional to the energy and to satisfy equation (11-6). In fact, it is necessary to have a distribution in which the points are more or less restricted to one part of the phase space, in order that the probability of phase can be sensibly used.

When it is desired to use a distribution in the phase space to represent a system of a definite energy but otherwise unknown initial conditions, it could be done by distributing the points over the surface that represents the desired energy. Then the average of the properties of the systems in the different parts of this surface might be thought to represent or at least to have something to do with the observable properties of a single system as it carries out its motion. Such a distribution, over a single energy surface, was named by Gibbs a *microcanonical ensemble.*

Although in thermodynamics one always talks about the energy of an isolated system, it is impossible to construct an isolated system. Any system is in contact with its surroundings and can exchange energy with them. The determination of the energy of a system by the means implied in the study of thermodynamics is a determination of some sort of average, or apparent, energy. On this account, the microcanonical ensemble does not exactly represent a thermodynamically isolated system. To get an ensemble that includes this possible variation of energy, Gibbs suggested and used the canonical ensemble. This is a distribution in which the probability of phase is given by

$$P = e^{\eta} = e^{(\psi - H)/\Theta} \tag{11-7}$$

In this expression H is the Hamiltonian function, and hence the energy, which is a function of position in the phase space. The quantities ψ and Θ are parameters characterizing the distribution. Θ is called the *modulus of the distribution,* and η is called the *index of probability.* ψ is given such a value that P satisfies the condition (11-6). Much of the study of statistical

mechanics consists in the study of the properties of this canonical ensemble and sometimes of the microcanonical ensemble.

4. The Fundamental Assumption of Statistical Mechanics as Applied to Thermodynamics.—Of course it is possible to study statistical mechanics for its own sake. It is possible to derive theorems concerning the behavior of distributions of points in the phase space for systems of various kinds, and these theorems will depend only upon the basic laws of mechanics. However, the object of most of the development of statistical mechanics has been the explanation of thermodynamic properties and the other properties of complex systems by means of the application of ordinary mechanics to the molecules of which the complex systems are composed. For this purpose it is necessary to know, or to assume, some connection between the ensemble whose properties are determined and the system whose properties it is desired to determine.

The fundamental assumption can be stated in various ways. Maxwell and Boltzmann made use of the ergodic hypothesis. According to this hypothesis, each representative point describes an orbit that eventually brings it into every portion of the phase space consistent with its energy. Thus any system, if left to itself long enough, will take on every possible configuration. In its extreme form this hypothesis is clearly untenable. The motion of an n-dimensional system is determined by $2n$ initial conditions, or $2n$ arbitrary constants. The values given these constants specify the orbits in the phase space, and different values give different orbits.

It seems much more satisfactory to recognize from the beginning the statistical nature of statistical mechanics. Only probability statements can be made, and these have nothing to do with sequences in time. Probabilities and average values can be evaluated, but the average refers to an average over different systems under the same macroscopic conditions, not to an average over the time for one system. Statistical mechanics is used because the information available about the system under consideration is incomplete. If nothing whatever is known about it, the reasonable assumption is that all positions of its

representative point in the phase space are equally probable. Such a system is well represented by a uniform distribution of points throughout the phase space. Such a distribution is constant in time, and the complete ignorance regarding the configuration of the system continues without change.

If the energy of the system is known, the representative point must lie on the surface corresponding to the given energy; but without additional information its location on the surface is entirely unknown. The proper representative ensemble is then one in which the density on the given energy surface is uniform and is elsewhere zero. Such a microcanonical ensemble is also in equilibrium and remains uniform over the energy surface.

If other constants of the motion are given, such as the momentum or the angular momentum, the uniform distribution is restricted to those regions of phase space corresponding to the specified quantities. In general, the rule for setting up a representative ensemble is that all parts of the phase space, consistent with the available information regarding the state of the system, are equally probable and must contain a uniform density of points. The probability that the system under consideration lies in a given volume of phase space consistent with its state is just proportional to the volume.

The above rule is easy to apply when certain constants of the motion are specified. It is less easy when thermodynamic quantities such as the temperature are given. However, various considerations suggest that a canonical ensemble is the proper representation of a system of which only the temperature is known. The modulus of the distribution Θ is proportional to the temperature, with a proportionality constant whose value can be determined.

According to the hypothesis just described, a problem in statistical mechanics consists first in setting up the proper ensemble to represent the available information (and lack of information) about the state of the system. The average values of various quantities can then be determined by averaging over the ensemble. The average value, over a canonical

ensemble, of any quantity u, which is a function of the canonical variables p_i and q_i, is given by the integral

$$\bar{u} = \int u e^{(\psi - H)/\Theta}\, dq_1 \cdots dq_n\, dp_1 \cdots dp_n \qquad (11\text{-}8)$$

It is also of interest to know whether all the values are close to the average or are widely distributed. This can be determined by evaluating the average square and higher powers. In many useful cases the spread of the distribution of values can be represented by the root-mean-square (rms) deviation.

$$(\Delta u)^2 = \overline{(u - \bar{u})^2} = \overline{u^2} - \bar{u}^2 \qquad (11\text{-}9)$$

To illustrate the procedure for averaging over a canonical ensemble consider a system that can be described by n normal coordinates q_i. The Hamiltonian function, or the energy, is given by

$$H = \sum_{i=1}^{n} (a_i p_i^2 + b_i q_i^2) \qquad (11\text{-}10)$$

and the average energy is

$$\bar{H} = \int \sum_{i=1}^{n} (a_i p_i^2 + b_i q_i^2) e^{\frac{\psi - \Sigma(a_i p_i^2 + b_i q_i^2)}{\Theta}} dq_1 \cdots dp_n \qquad (11\text{-}11)$$

To evaluate this integral consider first the average value of $a_1 p_1^2$ only.

$$\overline{a_1 p_1^2} = e^{\psi/\Theta} \int_{-\infty}^{\infty} a_1 p_1^2 e^{-a_1 p_1^2/\Theta}\, dp_1 \int_{-\infty}^{\infty} e^{-a_2 p_2^2/\Theta}\, dp_2$$

$$\cdots \int_{-\infty}^{\infty} e^{-b_n q_n^2/\Theta}\, dq_n = e^{\psi/\Theta} \left(\frac{\pi\Theta^3}{4a_1}\right)^{\frac{1}{2}} \prod_{i=2}^{n} \left(\frac{\pi\Theta}{a_i}\right)^{\frac{1}{2}} \prod_{i=1}^{n} \left(\frac{\pi\Theta}{b_i}\right)^{\frac{1}{2}}$$

To satisfy condition (11-5) $e^{-\psi/\Theta} = \prod_{i=1}^{n} \left(\frac{\pi\Theta}{a_i}\right)^{\frac{1}{2}} \left(\frac{\pi\Theta}{b_i}\right)^{\frac{1}{2}}$ so that

$$\overline{a_1 p_1^2} = \frac{\Theta}{2}$$

This is independent of the coefficient a_1 and is the same for all

coordinates and momenta. Hence

$$\bar{H} = n\Theta \tag{11-12}$$

This is the famous equipartition theorem according to which each degree of freedom in a system such as the one under consideration has the average energy Θ, half of it kinetic and half potential.

If the spread in a quantity is small compared with the magnitude of the average value of the quantity itself, the average value takes on added significance and may be regarded as the quantity to be expected under the given conditions. Hence the computation of average values and rms deviations can furnish an indication of the values of various quantities that are to be expected.

Problem 3. Prove the last equality in equation (11-9).

Problem 4. Show that in a canonical ensemble representing a system described by n normal coordinates

$$\Delta H = \left(\frac{2}{n}\right)^{\frac{1}{2}} \bar{H} \tag{11-13}$$

Problem 5. Consider an ensemble of systems each of which is a model of a perfect gas enclosed in a rectangular box. The molecules of each gas do not influence each other at all, and they have only kinetic energy. Show that the average energy of such an ensemble is equal to the number of molecules in a single gas multiplied by $3\Theta/2$. Show also that the rms deviation divided by the average energy is inversely proportional to the square root of the number of molecules in each body of gas.

Problem 6. Show that, if a mechanical system consists of two independent parts and if the system as a whole is represented by a canonical ensemble, the individual parts are distributed as a canonical distribution with the same modulus Θ.

Problem 7. By the use of the above problem to show that each molecule of a perfect gas can be represented by a canonical ensemble of molecules, find the density of a perfect gas as a function of position when it is subject to gravity.

5. Thermodynamic Analogies.—As has already been indicated, much of the interest in statistical mechanics lies in the light it sheds on the laws of thermodynamics. The behavior

of a thermodynamic system can be paralleled to a significant extent by the behavior of a canonical ensemble. This parallelism is based on a number of analogies.

a. The Internal Energy.—The above problems show that for a system composed of a large number of molecules, or having a large number of degrees of freedom, the average energy in a canonical ensemble closely represents the energy of all the systems. Hence the average energy of a canonical ensemble is taken as the analogue to the internal energy U of a system in thermodynamic equilibrium with its surroundings. The equality of the average energy of the ensemble and the internal energy of the thermodynamic system can be used as a criterion for the proper selection of the parameters ψ and Θ in order that the canonical ensemble shall properly represent the system.

b. The Temperature.—The modulus of distribution Θ shows a behavior similar to that of the temperature of a thermodynamic system. If two bodies have the same temperature and are put in contact, the composite system is in thermodynamic equilibrium at the same temperature unless there is a significant interaction between the two bodies. If there is a chemical reaction, for example, the resulting system will eventually come to equilibrium at a lower or a higher temperature than that of the individual systems before they were put in contact. If, however, the interaction is negligible, the whole system will be in equilibrium at the original temperature.

Each of the original systems can be represented by a canonical ensemble. When the two systems are put together, the probability of finding a representative point in a given element of phase space is equal to the product of the probabilities for the two portions of the phase space, or the two systems. The process is essentially that of combining each system of one ensemble with each system of the other to produce the composite ensemble. The probability of phase is then

$$P = e^{(\psi_1 - H_1)/\Theta} e^{(\psi_2 - H_2)/\Theta} = e^{(\psi_1 + \psi_2 - H)/\Theta} \qquad (11\text{-}14)$$

This is possible because only one value of Θ is involved and because in the absence of significant interactions $H_1 + H_2 = H$.

Although the quantity Θ shows a behavior analogous to that of temperature, it must be connected with the ordinary scale of temperature by a constant factor. Problem 5 leads to the result that for a perfect gas the average energy is $\frac{3}{2}n\Theta$, and from simple kinetic theory this energy is known to be $\frac{3}{2}nkT$, where n is the number of molecules in the gas and k is the molecular gas constant. Hence Θ may be identified with kT.

c. The Entropy.—The average value of the negative of the index of probability shows many analogies with the thermodynamic quantity entropy.

It can be proved that the average value of the index of probability for a canonical ensemble is less than for any other distribution with the same average energy. This can be shown as follows: Let η be the index of probability for the canonical distribution, *i.e.*, let $\eta = (\psi - H)/\Theta$, and let $\eta + \Delta\eta$ be any other value of the index leading to the same average energy. $\Delta\eta$ is then subject to two requirements. The definition of index of probability requires that

$$\int e^{\eta}\, dq_1 \cdots dq_n\, dp_1 \cdots dp_n = \int e^{\eta+\Delta\eta}\, dq_1 \cdots dq_n\, dp_1 \cdots dp_n = 1 \quad (11\text{-}15)$$

and the equality of the average energies provides the requirement that

$$\int He^{\eta}\, dq_1 \cdots dq_n\, dp_1 \cdots dp_n = \int He^{\eta+\Delta\eta}\, dq_1 \cdots dq_n\, dp_1 \cdots dp_n = \bar{H} \quad (11\text{-}16)$$

Then, because of these conditions,

$$\overline{\eta + \Delta\eta} - \bar{\eta} = \int [(\eta + \Delta\eta)e^{\eta+\Delta\eta} - \eta e^{\eta}]dq_1 \cdots dp_n$$

$$= \int \left[\frac{\psi - H}{\Theta}(e^{\eta+\Delta\eta} - e^{\eta}) + \Delta\eta e^{\eta+\Delta\eta}\right] dq_1 + \cdots dp_n$$

$$= \frac{\psi}{\Theta}\int (e^{\eta+\Delta\eta} - e^{\eta})dq_1 \cdots dp_n - \int \frac{H}{\Theta}(e^{\eta+\Delta\eta} - e^{\eta})dq_1 \cdots dp_n + \int \Delta\eta e^{\eta+\Delta\eta}\, dq_1 \cdots dp_n$$

The first integral in the last line vanishes because of equation

(11-15), and the second because of equation (11-16). Then, again because of equation (11-15),

$$\overline{\eta + \Delta\eta} - \bar{\eta} = \int \Delta\eta e^{\eta+\Delta\eta}\, dq_1 \cdots dp_n$$
$$= \int (\Delta\eta e^{\Delta\eta} + 1 - e^{\Delta\eta}) e^{\eta}\, dq_1 \cdots dp_n$$

The function in the parentheses is never negative so that its integral will be zero or positive, and $\bar{\eta}$ is a minimum under the specified conditions. The maximum nature of $(-\bar{\eta})$ is consistent with the idea of the increase in entropy as the system approaches equilibrium.

d. The Free Energy.—The quantity ψ, which is a constant and therefore does not have to be averaged over the ensemble, is analogous to the free energy F defined in equation (10-25). Since

$$\psi = \bar{H} + \Theta\eta \tag{11-17}$$

and Θ is taken as kT, this equation is analogous to

$$F = U - TS \tag{11-17a}$$

if η is taken as analogous to $-S/k$.

6. The Phase Integral.—From condition (11-6) it follows that

$$e^{-\psi/\Theta} = \int e^{-H/\Theta}\, dq_1 \cdots dp_n = f \tag{11-18}$$

The integral on the right is known as the *phase integral*, or the *partition function*. A number of averages can be evaluated by differentiating this phase integral, and the corresponding expressions are analogous to thermodynamic expressions.

The average energy can be obtained by differentiating with respect to Θ. Since Θ is not a function of the coordinates or momenta, the differentiation can be carried out under the integral sign. It follows directly that

$$\bar{H} = f^{-1}\Theta^2 \frac{\partial f}{\partial \Theta} = \Theta^2 \frac{\partial}{\partial \Theta} \log f = -\Theta^2 \frac{\partial}{\partial \Theta} (\psi/\Theta) \tag{11-19}$$

Upon inserting the corresponding thermodynamic quantities this becomes

$$U = F - T \frac{\partial F}{\partial T} \tag{11-20}$$

The phase integral depends on Θ and also on all the parameters of which H is a function. This may include such things as the positions of pistons or the strengths of electric or magnetic fields. If these parameters are designated by a_i and

$$A_i = -\frac{\partial H}{\partial a_i} \tag{11-21}$$

the average values of the A_i can be obtained by differentiating f. If a_i is a volume, A_i is a pressure. If an a_i is a magnetic field, A_i is a magnetic moment.

$$\bar{A}_i = -\frac{\overline{\partial H}}{\partial a_i} = \frac{\partial}{\partial a_i}\log f = -\frac{1}{\Theta}\frac{\partial \psi}{\partial a_i} \tag{11-22}$$

Since

$$\begin{aligned} \psi &= -\Theta \log f \\ d\psi &= -\Theta\frac{df}{f} - \log f\, d\Theta \\ &= -\frac{\Theta}{f}\left(\frac{\partial f}{\partial \Theta}\, d\Theta + \sum \frac{\partial f}{da_i}\, da_i\right) - \log f\, d\Theta \\ &= \left(\frac{\psi}{\Theta} - \frac{U}{\Theta}\right) d\Theta - \sum A_i\, da_i \end{aligned} \tag{11-23}$$

This is analogous to equation (10-20) with $\sum_i A_i\, da_i$ as a generalization of $p\, dv$.

A canonical ensemble represents a system in a certain thermodynamic state. The independent variables that define the state are sufficient to define the corresponding ensemble, by means of the analogies just discussed. Other variables of state are the volume, which appears in the statistical treatment as one of the external parameters a_i, and the pressure, which is the corresponding A_i. When the proper ensemble has been determined, the fundamental postulate can be applied to obtain a statistical interpretation of the thermodynamic quantities.

Problem 8. Show that, when the external parameters are held constant,

$$\frac{\partial \psi}{\partial \Theta} = \bar{\eta} \tag{11-24}$$

and show that this is analogous to the corresponding equation of thermodynamics.

Problem 9. Show that

$$\bar{H}^s = e^{\psi/\Theta}\left(\Theta^2 \frac{\partial}{\partial\Theta}\right)^s e^{-\psi/\Theta} \tag{11-25}$$

References

FOWLER, R. H.: "Statistical Mechanics," Cambridge University Press, London.

GIBBS, J. WILLARD: "Elementary Principles in Statistical Mechanics"; or "Collected Works," Vol. II, Longmans, Green & Co., Inc., New York.

LORENTZ, H. A.: "Lectures on Theoretical Physics," Vol. II, Macmillan & Co., Ltd., London.

TOLMAN, R. C.: "The Principles of Statistical Mechanics," Oxford University Press, London.

CHAPTER XII

THE VECTOR FIELD

If with every point in space there is associated a certain vector, the whole set of vectors is said to constitute a *vector field.* The association of the vectors with the points in space is the same as the association of the values of a function with the values of the independent variables. In fact, the vector is a function of the three coordinates of position. Since a vector has three components, a vector field can be represented by three independent functions of the three independent coordinates of position. The linear vector function described in Chap. VIII is a special case of a vector field if the independent vector is regarded as giving the location of a point in space. The general vector field is not limited to linear functions; but, in most physically significant cases, each component of the vector field can be differentiated with respect to the three independent variables x, y, z.

A field of force, such as the earth's gravitational field or the electric field around a charged conductor, is a vector field. For every point there is a vector representing the force on a particle placed at the point. Each component of the force is a scalar function of the coordinates.

1. The Gradient.—A type of vector field which is most useful in physical problems is one in which the vector can be derived from a scalar function of position by differentiation. If there exists a single-valued differentiable function $V(x,y,z)$, the gradient of V is defined by the equation

$$\textbf{grad } V = \mathbf{i}\frac{\partial V}{\partial x} + \mathbf{j}\frac{\partial V}{\partial y} + \mathbf{k}\frac{\partial V}{\partial z} \tag{12-1}$$

Thus **grad** V is a vector, even though V itself is a scalar. Since V is a function of three independent variables, the partial deriva-

tive signs must be used, and the components of the gradient are the partial derivatives of the scalar along the three coordinate axes.

The points at which the function V has a given value C will lie on the surface $V(x,y,z) = C$. Then let $\mathbf{i}\,dx + \mathbf{j}\,dy + \mathbf{k}\,dz$ be an infinitesimal vector lying in this surface. The statement that this vector lies in the surface implies a relationship between dx, dy, and dz. This relationship can be obtained by differentiating the equation of the surface.

$$dV = \frac{\partial V}{\partial x}\,dx + \frac{\partial V}{\partial y}\,dy + \frac{\partial V}{\partial z}\,dz = 0 \tag{12-2}$$

If dx and dy are arbitrarily chosen, dz can be determined from equation (12-2) so the vector will lie in the surface.

$$dz = \frac{-\left(\dfrac{\partial V}{\partial x}\,dx + \dfrac{\partial V}{\partial y}\,dy\right)}{\dfrac{\partial V}{\partial z}} \tag{12-3}$$

The scalar product of **grad** V and the vector in the surface is then

$$\frac{\partial V}{\partial x}\,dx + \frac{\partial V}{\partial y}\,dy + \frac{\partial V}{\partial z}\,dz = 0 \tag{12-4}$$

so that the gradient is perpendicular to every vector lying in the surface $V = C$. The gradient represents the direction and magnitude of the rate of change of V normal to the surface on which V is constant and thus the direction and magnitude of the greatest rate of increase of V.

Problem 1. Find the gradient of the potential energy of a particle near the surface of the earth. Take first a system of coordinates in which one axis is vertical. Then take a system of axes in which one axis is parallel to the surface while the other axes make angles of 45° with the vertical.

Problem 2. Find the gradient of V when $V = (x^2 + y^2)$, and show that this represents the direction and magnitude of the greatest rate of increase of this function.

Problem 3. Show that the increase in V due to a change in position **dr** is given by $dV = \mathbf{dr} \cdot \mathbf{grad}\ V$.

Problem 4. Find the gradient of a/r, where a is a constant and r is the length of the vector from the origin to the point in question.

Problem 5. Find the gradient of the potential energy of a particle attracted toward the origin with a force proportional to the distance from it.

Problem 6. Show that the force on a particle is given by the negative gradient of its potential energy.

Problem 7. Show that the components of the gradient transform properly for them to be the components of a vector.

Problem 8. Show that **grad** $(\mathbf{w} \cdot \mathbf{r}) = \mathbf{w}$, where $\mathbf{w}$ is a constant vector and $\mathbf{r}$ is the radius vector from the origin.

2. The Divergence.—If $\mathbf{a} = a_x\mathbf{i} + a_y\mathbf{j} + a_z\mathbf{k}$, the divergence of $\mathbf{a}$ is the scalar quantity defined by the equation

$$div\ \mathbf{a} = \frac{\partial a_x}{\partial x} + \frac{\partial a_y}{\partial y} + \frac{\partial a_z}{\partial z} \tag{12-5}$$

The divergence has no meaning, of course, for an isolated vector. It can be formed only in a vector field, where each component of the vector is a function of the three coordinates x, y, and z. It is necessary to distinguish clearly between the components of the dependent vector, a_x, a_y, and a_z, and the coordinates of position x, y, and z. The coordinates of position represent the components of the vector from the origin to the point with which the dependent vector is associated and therefore may be considered as the components of the independent vector; usually, however, they are considered as merely the coordinates of this "field point."

The divergence gets its name from association with the idea of the flow of a liquid. Suppose a vector $\mathbf{a}$ represents the velocity of flow at each point in a liquid. Then consider a small element of volume $dx\ dy\ dz$, the coordinates of whose center are x, y, and z. The volumes of liquid that flow in unit time through the two faces perpendicular to the x axis are, respectively, $\left(a_x - \frac{1}{2}\frac{\partial a_x}{\partial x}\,dx\right)dy\,dz$ and $\left(a_x + \frac{1}{2}\frac{\partial a_x}{\partial x}\,dx\right)dy\,dz$. The excess of the volume flowing out over the volume flowing in is $(\partial a_x/\partial x)dx\,dy\,dz$. Similar considerations for the other two pairs of faces show that the total excess of the volume flowing

out of the element over the volume flowing into it is

$$\left(\frac{\partial a_x}{\partial x} + \frac{\partial a_y}{\partial y} + \frac{\partial a_z}{\partial z}\right) dx\, dy\, dz = div\ \mathbf{a}\, dv$$

The divergence of the velocity is thus the volume of liquid that diverges from a unit volume per unit of time.

It is often convenient to use the symbolic operator ∇ to represent the differential operations in connection with vector fields. This operator is defined by

$$\nabla = \mathbf{i}\frac{\partial}{\partial x} + \mathbf{j}\frac{\partial}{\partial y} + \mathbf{k}\frac{\partial}{\partial z} \tag{12-6}$$

Problem 9. Find the divergence of $\mathbf{r}$, where $\mathbf{r}$ is the vector from the origin to the point with which it is associated. In this problem care must be taken to distinguish between x, y, and z as components of the dependent variable $\mathbf{r}$ and x, y, and z as components of the independent vector of position.

Problem 10. Find *div* **grad** V, where V is any differentiable scalar function of position.

Problem 11. Show that the divergence is invariant under a rotation of coordinates.

Problem 12. If the scalar S and the vector $\mathbf{A}$ are both functions of position, show that

$$div\ (S\mathbf{A}) = S\ div\ \mathbf{A} + \mathbf{A} \cdot \mathbf{grad}\ S \tag{12-7}$$

3. The Curl.—The **curl a,** or $(\nabla \times \mathbf{a})$, is a vector and is defined by

$$\mathbf{curl\ a} = \nabla \times \mathbf{a} = \mathbf{i}\left(\frac{\partial a_z}{\partial y} - \frac{\partial a_y}{\partial z}\right) + \mathbf{j}\left(\frac{\partial a_x}{\partial z} - \frac{\partial a_z}{\partial x}\right) + \mathbf{k}\left(\frac{\partial a_y}{\partial x} - \frac{\partial a_x}{\partial y}\right) \tag{12-8}$$

The form of this definition can be remembered by writing it as the determinant for the vector product. In this sense it is the vector product of the vector operator ∇ and $\mathbf{a}$.

$$\nabla \times \mathbf{a} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ a_x & a_y & a_z \end{vmatrix} . \tag{12-9}$$

The curl is of especial use in the study of rotational motion, for it relates the linear velocity of the particles of a body to the angular velocity of the body as a whole. If a solid is rotating about an axis, twice the angular velocity of this rotation is equal to the curl of the linear velocity of the particles of the body. The curl is also of use in electromagnetic theory, since it relates a magnetic field to the current that produces it.

Problem 13. A solid body is turning about a fixed axis with a constant angular velocity $\boldsymbol{\omega}$. Find the linear velocity of the particles of the body, and find the curl of these velocities in terms of $\boldsymbol{\omega}$.

Problem 14. Show that

$$\textbf{curl grad } V = \nabla \times \nabla V = 0$$

This is a very important property of the gradient and shows that, unless the curl of a vector field is everywhere zero, the vector cannot be the gradient of a scalar.

Problem 15. Evaluate *div* **curl a** in terms of the Cartesian components.

Problem 16. Show that

$$div\ (\mathbf{A} \times \mathbf{B}) = \mathbf{B} \cdot \textbf{curl A} - \mathbf{A} \cdot \textbf{curl B} \qquad (12\text{-}10)$$

4. The Line Integral.—The line integral of a vector **a** along a curve L is written $\int_L \mathbf{a} \cdot \mathbf{dl}$. It is the integral of the scalar product of the vector **a** and the vector element of the curve **dl**. The notation implies that the vector **a** is a function of position, and its value at the location of **dl** is to be taken in forming the scalar product. If **a** represents the force on a particle as a function of its position and L is the path along which it moves, the line integral is the amount of work done by the force during the motion. It is thus useful in evaluating potential energies.

The line integral can be written in terms of the components of **a** and **dl** or the components of **a** and the direction cosines λ, μ, ν of the element **dl**.

$$\mathbf{a} \cdot \mathbf{dl} = (\lambda a_x + \mu a_y + \nu a_z)dl = (a_x\,dx + a_y\,dy + a_z\,dz) \qquad (12\text{-}11)$$

Problem 17. Show that the line integral of a constant vector around any closed curve is equal to zero.

Problem 18. Consider the vector field $\mathbf{a} = y\mathbf{i}$, and evaluate the line integral of $\mathbf{a}$ around a square in the x-y plane.

Problem 19. Evaluate the line integral of $\mathbf{a} = y\mathbf{i}$ around a circular path in the x-y plane.

Problem 20. Show that the line integral $\int_L \mathbf{grad}\ V \cdot \mathbf{dl}$ between any two points is independent of the path and hence that this integral taken around any closed curve is zero.

Problem 21. A particle moves in an ellipse under the influence of an inverse-square attraction toward one focus. Find the work done on the particle as it moves from one end of the major axis to the other. Also find the work done while it moves from one end of the minor axis to the other.

Problem 22. Evaluate the line integral of $\mathbf{a} = x^2\mathbf{i}$ around a semicircle joining the origin with the point $(x = b, y = 0)$. Also evaluate the integral along a straight line from the origin to $(x = b/2, y = c)$ and then along another line to $(x = b, y = 0)$.

Problem 23. Evaluate the line integral of $\mathbf{a} = y^2\mathbf{i} - x^2\mathbf{j}$ along the paths in the above problem.

5. The Surface Integral.—It is often convenient to represent a plane surface by a vector, perpendicular to the surface, whose length is proportional to the area. Some convention must be adopted as to the positive direction of such a vector. If the surface encloses a volume, the positive direction is taken outward. If a line integral is taken around the surface, the direction in which the boundary is traversed is related to the positive direction of the surface vector by the right-handed-screw rule.

A curved surface may be considered as the sum of infinitesimal plane surfaces in the way that a curved line is regarded as made up of infinitesimal straight elements. The surface integral of a vector $\mathbf{a}$ over a surface S is defined as the integral of the scalar product of the vector $\mathbf{a}$ and the infinitesimal vector $\mathbf{ds}$ that represents the surface element. The surface integral is written $\int_S \mathbf{a} \cdot \mathbf{ds}$. Since $\mathbf{ds}$ is an outward-drawn normal to the surface, the surface integral is the integral of the normal component of the vector $\mathbf{a}$. Evidently the surface integral of a constant vector over a plane surface is just the scalar product of the constant vector and the vector representing the surface. If the vector $\mathbf{a}$ represents the velocity of a fluid, the surface

integral of **a** over a surface represents the volume of the fluid that flows through the surface per unit time.

Problem 24. Evaluate the surface integral of a constant vector $\mathbf{a} = a\mathbf{i}$ over the surface of a sphere.

Problem 25. Show that $\int_S \mathbf{A} \cdot \mathbf{ds} = 0$, when **A** is a constant vector and S is a closed surface.

Problem 26. Find the value of $\int_S \mathbf{r} \cdot \mathbf{ds}$, when S is a sphere whose center is at the origin and **r** is the radius vector from the center.

Problem 27. Show that the volume enclosed by any closed surface is given by

$$3V = \int_S \mathbf{r} \cdot \mathbf{ds} \tag{12-12}$$

6. Gauss's Theorem.—An important theorem in potential theory is known as *Gauss's theorem.* It states that the surface integral of a vector over a closed surface is equal to the volume integral of the divergence of the vector throughout the enclosed volume. In the usual notation

$$\int_V \nabla \cdot \mathbf{A}\, dv = \int_S \mathbf{A} \cdot \mathbf{ds} \tag{12-13}$$

By means of this theorem it is often possible to transform a surface integral into a volume integral or a volume integral into a surface integral. The physical significance of the theorem is very simple in some cases. If the vector **A** represents the velocity of flow of a fluid, the integral of the divergence is the volume of material created inside the surface, while the surface integral is the amount that flows out through the surface per unit time.

An elementary proof of this theorem can be given as follows: If equation (12-13) is written in terms of its components, it becomes

$$\begin{aligned}\int_V \nabla \cdot \mathbf{A}\, dv &= \int \frac{\partial A_x}{\partial x}\, dx\, dy\, dz + \int \frac{\partial A_y}{\partial y}\, dx\, dy\, dz \\ &\qquad + \int \frac{\partial A_z}{\partial z}\, dx\, dy\, dz \\ &= \int A_x\, dy\, dz + \int A_y\, dx\, dz + \int A_z\, dx\, dy \\ &= \int_S \mathbf{A} \cdot \mathbf{ds}\end{aligned}$$

The sequence of the reasoning can be followed with the aid of Fig. 12-1. Consider first the integration of the term containing A_x, and integrate this with respect to x. The result is the value of A_x at the upper limit of x minus the corresponding value at the lower limit of x. The figure shows a rectangular prism of cross section $dy\ dz$ cut out of the volume over which the integration is to be carried out. $\mathbf{ds}_1$ is the element of surface cut out by this prism on the right-hand side of the volume, while

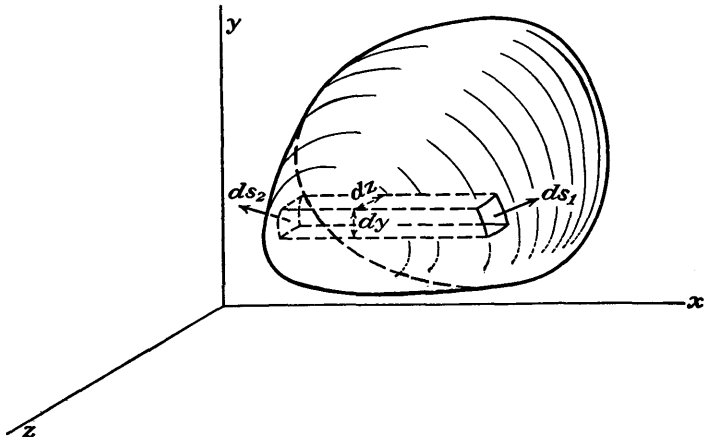


FIG. 12-1.—Transformation of a volume integral into a surface integral by means of Gauss's theorem.

$\mathbf{ds}_2$ is the corresponding surface element on the other side. $dy\ dz$ is the projection of $\mathbf{ds}_1$ perpendicular to the x axis, while $-dy\ dz$ is the corresponding projection of $\mathbf{ds}_2$. These projections multiplied by the values of A_x at the corresponding positions are those elements of the surface integral which are contributed by the elements of surface $\mathbf{ds}_1$ and $\mathbf{ds}_2$. Hence the integration with respect to x gives an element of the surface integral, and the integration with respect to y and z gives the complete surface integral of the x component of $\mathbf{A}$ multiplied by the x component of the vector representing the surface elements. A similar procedure with A_y and A_z gives the complete surface integral. It will be noted that the elements of the surface used in carrying out the integration of A_y are not the same as those used for A_x. However, all three integrations cover the entire surface.

Problem 28. Show what additions must be made to the above proof of Gauss's theorem when the surface has some reentrant portions.

Problem 29. Show that the volume enclosed by a closed surface is given by

$$V = \tfrac{1}{6} \int_S \mathbf{grad}\ (r^2) \cdot \mathbf{ds}$$

where $\mathbf{r}$ is the radius vector from the origin.

Problem 30. Show that

$$\int_V B\ div\ \mathbf{A}\ dv + \int_V \mathbf{A} \cdot \mathbf{grad}\ B\ \ dv = \int_S B\mathbf{A} \cdot \mathbf{ds} \qquad (12\text{-}14)$$

where the surface S surrounds the volume V.

7. Stokes's Theorem.—Stokes's theorem states that the surface integral of the curl of a vector over any surface is equal to the line integral of the vector around the boundary of the surface.

$$\int_S \mathbf{curl}\ \mathbf{A} \cdot \mathbf{ds} = \int_L \mathbf{A} \cdot \mathbf{dl} \qquad (12\text{-}15)$$

This is analogous to Gauss's theorem in a smaller number of dimensions, since it is a relation between a surface integral and a line integral.

To prove this theorem consider a square element $dx\ dy$ of the surface S, and for convenience consider the origin of coordinates as at the center of the element. Let the components of the vector $\mathbf{A}$ at the origin be A_x, A_y, and A_z. Then evaluate the line integral of the vector around this element.

$$\begin{aligned}\mathbf{A} \cdot \mathbf{dl} &= \left(A_x - \frac{1}{2}\frac{\partial A_x}{\partial y}\,dy\right)dx + \left(A_y + \frac{1}{2}\frac{\partial A_y}{\partial x}\,dx\right)dy \\ &\quad - \left(A_x + \frac{1}{2}\frac{\partial A_x}{\partial y}\,dy\right)dx - \left(A_y - \frac{1}{2}\frac{\partial A_y}{\partial x}\,dx\right)dy \\ &= \left(\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}\right)dx\,dy = \mathbf{curl}\,\mathbf{A} \cdot \mathbf{ds}\end{aligned}$$

This relation for the single surface element is clearly independent of the orientation or the location of the element and thus will hold for every square element of the surface. Since every surface can be approximated as closely as is desired by square elements and since the line integral around the whole surface is

equal to the sum of the line integrals about any parts into which it may be divided, the theorem is proved.

Problem 31. Show by a diagram that the line integral about a surface is equal to the sum of the line integrals about any parts into which it may be divided.

Problem 32. Illustrate Stokes's theorem by evaluating the line integral, and the surface integral of the curl, of the vector $\mathbf{r} \times \mathbf{k}$, where $\mathbf{k}$ is the unit vector along the z axis and $\mathbf{r}$ is the radius vector from the origin. Take as the surface a hemisphere the center of which is at the origin and which is bounded by the x-y plane.

Problem 33. Show that, if $\nabla \times \mathbf{A} = 0$, the line integral of $\mathbf{A}$ between any two points is independent of the path.

Problem 34. Show that

$$\int_V \text{curl } \mathbf{A}\, dv = -\int_S \mathbf{A} \times \mathbf{ds} \tag{12-16}$$

8. Tensor Fields.—A vector is a special case of a tensor, and a vector field is therefore a special case of a tensor field. If each component of a tensor is a function of the three coordinates of position, various differential operations can be carried out and the results of these operations are again tensors. All the considerations given in this section refer directly to tensors whose components are with reference to Cartesian axes. Care must be exercised in extending them to other systems.

a. The Gradient of a Tensor.—As already indicated, the gradient of a scalar is a vector. Since the scalar is a tensor of rank 0 and the vector is a tensor of rank 1, the operation of taking the gradient increases by 1 the rank of the tensor. A similar operation on a vector, or tensor of rank 1, leads to a tensor of rank 2, $\partial a_i/\partial x_j$. The fact that this is a tensor follows from the way in which it will transform when the coordinate axes are rotated about the origin.

A tensor of rank 2 may be regarded as the sum of a symmetric and an antisymmetric tensor.

$$\frac{\partial a_i}{\partial x_j} = \frac{1}{2}\left(\frac{\partial a_i}{\partial x_j} + \frac{\partial a_j}{\partial x_i}\right) + \frac{1}{2}\left(\frac{\partial a_i}{\partial x_j} - \frac{\partial a_j}{\partial x_i}\right) \tag{12-17}$$

The antisymmetric tensor in equation (12-17) is what has been called the curl and considered to be a vector. This identifica-

tion of an antisymmetric tensor of the second rank with a vector is possible only in three dimensions. If the vector a_i is the gradient of a scalar, the antisymmetric tensor vanishes and the whole derivative tensor is symmetric.

An important property of the gradient of a scalar is that its line integral between two points is independent of the path. The gradient of a vector has somewhat similar properties.

$$\int_a^b \frac{\partial a_i}{\partial x_j} dx_j = a_i \Big|_a^b \tag{12-18}$$

and the line integral is again independent of the path.

b. The Divergence.—While taking the gradient increases the rank of a tensor by 1, taking the divergence decreases it by the same amount. The divergence of a vector is a scalar, and the divergence of a tensor of rank 2 is a vector, or tensor of rank 1.

$$(div\ T)_k = \frac{\partial T_{ik}}{\partial x_i} \tag{12-19}$$

If the tensor T is symmetric,

$$\frac{\partial T_{ik}}{\partial x_i} = \frac{\partial T_{ki}}{\partial x_i}$$

so that the definition of the divergence does not require specification of whether the derivative is with respect to the first or second index.

A theorem very similar to Gauss's theorem holds for the divergence of a symmetric tensor.

$$\int_V \frac{\partial T_{ik}}{\partial x_i} dv = \int_S T_{ik}\, ds_k \tag{12-20}$$

The proof of this theorem can be carried through just as the proof of Gauss's theorem on page 237 and need not be given in detail here.

9. Orthogonal Curvilinear Coordinates.—In the previous sections it has been assumed that positions in space were located in Cartesian coordinates and also that the vectors were described in the same set of coordinates. This is not always convenient, and therefore it is desirable to derive the expressions for the various differential operations when other coordinates are used.

A three-dimensional coordinate system consists of three

sets of surfaces that intersect each other. On each surface a certain quantity, the coordinate, is constant. This coordinate has a different value for each surface of the set, and it will be assumed that there is a continuum of surfaces, represented by all possible values of the coordinate. In Cartesian coordinates the surfaces are planes intersecting each other at right angles. The intersection of three of these planes determines a point, and the point is then designated by the values of the three coordinates that specify the planes. Similarly, in spherical polar coordinates the surfaces are a set of concentric spheres specified by the values of r, a set of planes which all pass through the polar axis and are specified by the values of φ, and a set of circular cones of which the vertices are all at the origin and which are specified by the values of the variable θ. In these two examples the surfaces intersect each other at right angles, and consequently these coordinates are examples of orthogonal coordinates. Only orthogonal coordinates will be considered here.

Let the three coordinates be q_1, q_2, and q_3. Because of the orthogonality of the coordinate surfaces, it is possible to set up, at any point, an orthogonal set of unit vectors $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$, in the directions of increasing q_1, q_2, q_3, respectively. It is important to select q_1, q_2, and q_3, so that the unit vectors $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ form a right-hand system of axes. This set of three vectors defines a Cartesian coordinate system that coincides with the curvilinear system in the immediate neighborhood of this one point. The differentials of these systems of coordinates are connected with the differentials of the q's by the relations

$$ds_1 = h_1\,dq_1 \qquad ds_2 = h_2\,dq_2 \qquad ds_3 = h_3\,dq_3 \quad (12\text{-}21)$$

In cylindrical coordinates

$$ds_1 = dr \qquad ds_2 = r\,d\theta \qquad ds_3 = dz$$

so that

$$h_1 = 1 \qquad h_2 = r \qquad h_3 = 1 \quad (12\text{-}22)$$

The quantities h_1, h_2, and h_3 are functions of q_1, q_2, q_3 and vary from point to point. In general, the q's do not have the dimensions of length, and the quantities h are necessary to translate a change in q into a length. The general expression for the

gradient is

$$\textbf{grad } V = \frac{1}{h_1}\frac{\partial V}{\partial q_1}\mathbf{e}_1 + \frac{1}{h_2}\frac{\partial V}{\partial q_2}\mathbf{e}_2 + \frac{1}{h_3}\frac{\partial V}{\partial q_3}\mathbf{e}_3 \qquad (12\text{-}23)$$

This includes the expression (12-1) for the case of Cartesian coordinates, since then $h_1 = h_2 = h_3 = 1$ and $q_1 = x$, etc. It is important also to notice that the unit vectors are not necessarily constant but may vary in direction from point to point.

By similar analysis it is possible to derive general expressions for other functions in vector fields, but in doing so account must be taken of the variation in direction from point to point of the unit vectors. However, probably the simplest derivation of the expression for the divergence is not directly, but by means of Gauss's theorem. Consider the volume element bounded by the surfaces

$$q_1 + \frac{dq_1}{2}, \quad q_1 - \frac{dq_1}{2}, \quad q_2 + \frac{dq_2}{2}, \quad q_2 - \frac{dq_2}{2}, \quad q_3 + \frac{dq_3}{2}, \quad q_3 - \frac{dq_3}{2}$$

The center point is at $q_1q_2q_3$. Neglecting small quantities of higher order, the volume of this element is $h_1h_2h_3\,dq_1\,dq_2\,dq_3$, and the divergence at the center multiplied by the volume will be equal to the integral of the vector over the bounding surface.

Consider first the two surfaces to which $\mathbf{e}_1$ is perpendicular. These surfaces differ in area because h_2 and h_3 are changing with q_1. If $\mathbf{A}$ is the vector whose divergence is to be found,

$$A_1 = \mathbf{A} \cdot \mathbf{e}_1$$

also changes with q. Hence the integral over one face is $-\left[A_1h_2h_3 - \frac{\partial}{\partial q_1}(A_1h_2h_3)\frac{dq_1}{2}\right]dq_2\,dq_3$ and over the other face is $\left[A_1h_2h_3 + \frac{\partial}{\partial q_1}(A_1h_2h_3)\frac{dq_1}{2}\right]dq_2\,dq_3$. The integrals over the other four faces can be similarly expressed so that

$$\begin{aligned} &div\ \mathbf{A}h_1h_2h_3\,dq_1\,dq_2\,dq_3 \\ &\quad = \left[\frac{\partial}{\partial q_1}(A_1h_2h_3) + \frac{\partial}{\partial q_2}(A_2h_1h_3) + \frac{\partial}{\partial q_3}(A_3h_1h_2)\right]dq_1\,dq_2\,dq_3 \end{aligned}$$

Hence

$$div\ \mathbf{A} = \frac{1}{h_1h_2h_3}\left[\frac{\partial}{\partial q_1}(A_1h_2h_3) + \frac{\partial}{\partial q_2}(A_2h_1h_3) + \frac{\partial}{\partial q_3}(A_3h_1h_2)\right] \qquad (12\text{-}24)$$

Problem 35. Show that the expression for the gradient in cylindrical coordinates is

$$\textbf{grad } V = \mathbf{e}_r \frac{\partial V}{\partial r} + \mathbf{e}_\theta \frac{\partial V}{r\,\partial \theta} + \mathbf{e}_z \frac{\partial V}{\partial z} \tag{12-25}$$

Problem 36. Show that the expression for the gradient in spherical polar coordinates is

$$\textbf{grad } V = \mathbf{e}_r \frac{\partial V}{\partial r} + \mathbf{e}_\theta \frac{\partial V}{r\,\partial \theta} + \mathbf{e}_\varphi \frac{1}{r \sin\theta} \frac{\partial V}{\partial \varphi} \tag{12-26}$$

Problem 37. Show that the expression for the divergence in spherical polar coordinates is

$$div\ \mathbf{A} = \frac{1}{r^2}\frac{\partial}{\partial r}(r^2 A_r) + \frac{1}{r \sin\theta}\frac{\partial}{\partial \theta}(\sin\theta\, A_\theta) + \frac{1}{r \sin\theta}\frac{\partial A_\varphi}{\partial \varphi} \tag{12-27}$$

Problem 38. Show that in orthogonal curvilinear coordinates,

$$div\ \textbf{grad } V = \frac{1}{h_1 h_2 h_3}\left[\frac{\partial}{\partial q_1}\left(\frac{h_2 h_3}{h_1}\frac{\partial V}{\partial q_1}\right) + \frac{\partial}{\partial q_2}\left(\frac{h_1 h_2}{h_2}\frac{\partial V}{\partial q_2}\right) + \frac{\partial}{\partial q_3}\left(\frac{h_1 h_2}{h_3}\frac{\partial V}{\partial q_3}\right)\right] \tag{12-28}$$

Problem 39. Show that in spherical polar coordinates

$$div\ \textbf{grad } V = \frac{1}{r^2}\left[\frac{\partial}{\partial r}\left(r^2\frac{\partial V}{\partial r}\right) + \frac{1}{\sin\theta}\frac{\partial}{\partial \theta}\left(\sin\theta\frac{\partial V}{\partial \theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2 V}{\partial \varphi^2}\right] \tag{12-29}$$

Problem 40. Show that

$$h_i^2 = \left(\frac{\partial x}{\partial q_i}\right)^2 + \left(\frac{\partial y}{\partial q_i}\right)^2 + \left(\frac{\partial z}{\partial q_i}\right)^2 \tag{12-30}$$

where x, y, and z are fixed Cartesian coordinates.

Problem 41. The expression for the curl in curvilinear coordinates can be obtained by applying Stokes's theorem to an elementary rectangle perpendicular to each of the unit vectors, much as the above expression for the divergence was obtained. In this way show that

$$\begin{aligned}\textbf{curl } \mathbf{A} = {} & \frac{\mathbf{e}_1}{h_2 h_3}\left[\frac{\partial}{\partial q_2}(h_3 A_3) - \frac{\partial}{\partial q_3}(h_2 A_2)\right] \\ & + \frac{\mathbf{e}_2}{h_1 h_3}\left[\frac{\partial}{\partial q_3}(h_1 A_1) - \frac{\partial}{\partial q_1}(h_3 A_3)\right] \\ & + \frac{\mathbf{e}_3}{h_1 h_2}\left[\frac{\partial}{\partial q_1}(h_2 A_2) - \frac{\partial}{\partial q_2}(h_1 A_1)\right]\end{aligned} \tag{12-31}$$

Problem 42. Show that in spherical polar coordinates

$$\mathbf{curl\ A} = \frac{\mathbf{e}_r}{r \sin\theta}\left[\frac{\partial}{\partial\theta}(\sin\theta A_\varphi) - \frac{\partial A_\theta}{\partial\varphi}\right] + \frac{\mathbf{e}_\theta}{r}\frac{1}{\sin\theta}\left[\frac{\partial A_r}{\partial\varphi} - \frac{\partial}{\partial r}(r\sin\theta A_\varphi)\right] + \frac{\mathbf{e}_\varphi}{r}\left[\frac{\partial}{\partial r}(rA_\theta) - \frac{\partial A_r}{\partial\theta}\right] \quad (12\text{-}32)$$

10. Vector Identities.—There are several identical relationships between various functions of a vector field that are of use in the treatment of physical problems. Some of these are tabulated here for future reference. Several have already been proved, while the proofs of the others can be carried out by means of the definitions of the various operations.

(1) $\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = \mathbf{B}(\mathbf{A}\cdot\mathbf{C}) - \mathbf{C}(\mathbf{A}\cdot\mathbf{B})$

(2) $div\ \mathbf{curl\ A} = 0$

(3) $div\ (S\mathbf{A}) = S\ div\ \mathbf{A} + \mathbf{A}\cdot\mathbf{grad}\ S$

(4) $div\ (\mathbf{A}\times\mathbf{B}) = \mathbf{B}\cdot\mathbf{curl\ A} - \mathbf{A}\cdot\mathbf{curl\ B}$

(5) $\mathbf{curl\ grad}\ S = 0$

(6) $\mathbf{curl}\ (S\mathbf{A}) = S\ \mathbf{curl\ A} + \mathbf{grad}\ S \times \mathbf{A}$

(7) $\mathbf{curl\ curl\ A} = \mathbf{grad}\ div\ \mathbf{A} - \nabla^2\mathbf{A}$

(8) $\mathbf{curl}\ (\mathbf{A}\times\mathbf{B}) = \mathbf{A}\ div\ \mathbf{B} - \mathbf{B}\ div\ \mathbf{A} + (\mathbf{B}\cdot\nabla)\mathbf{A} - (\mathbf{A}\cdot\nabla)\mathbf{B}$

(9) $\mathbf{grad}\ (\mathbf{A}\cdot\mathbf{B}) = (\mathbf{A}\cdot\nabla)\mathbf{B} + (\mathbf{B}\cdot\nabla)\mathbf{A} + \mathbf{A}\times\mathbf{curl\ B} + \mathbf{B}\times\mathbf{curl\ A}$

These identities are independent of the coordinate system used, but the expressions $\nabla^2\mathbf{A}$, where ∇^2 is applied to a vector, and $(\mathbf{A}\cdot\nabla)\mathbf{B}$ have not yet been defined. In Cartesian coordinates

$$\nabla^2\mathbf{A} = \mathbf{i}(\nabla^2 A_x) + \mathbf{j}(\nabla^2 A_y) + \mathbf{k}(\nabla^2 A_z)$$

and

$$(\mathbf{A}\cdot\nabla)\mathbf{B} = \mathbf{i}\left(A_x\frac{\partial B_x}{\partial x} + A_y\frac{\partial B_x}{\partial y} + A_z\frac{\partial B_x}{\partial z}\right) + \mathbf{j}\left(A_x\frac{\partial B_y}{\partial x} + A_y\frac{\partial B_y}{\partial y} + A_z\frac{\partial B_y}{\partial z}\right) + \mathbf{k}\left(A_x\frac{\partial B_z}{\partial x} + A_y\frac{\partial B_z}{\partial y} + A_z\frac{\partial B_z}{\partial z}\right)$$

In other systems of coordinates these must be properly transformed.

11. The Potential.—As was pointed out in Chap. II, if the force on a particle can be expressed as the negative gradient of a scalar function of the particle's position, this scalar function

may be called the potential energy of the particle in the given field of force. The use of a potential energy is a very convenient method of representing forces due to a number of different attracting centers, for the potential energies due to the different centers can be merely added together. It is often easier to add the scalar potentials than to add the vector forces.

From the solution of Prob. 4, it follows that the potential energy of a system consisting of an element of matter fixed at the origin and a unit mass at the end of the vector r is

$$dV = -\frac{K\rho\, dv}{r} \tag{12-33}$$

where ρ is the density and dv is the volume of the element at the origin. This potential energy with respect to a unit mass at the end of r may be called the potential at the end of r due to the element of mass at the origin. A similar terminology is used in electrostatics, where a unit positive charge is used instead of a unit mass. If the attracting body is not small enough to be considered as a single particle, the potential due to the whole is merely the sum of the potentials due to the individual elements.

$$V = -K \int \frac{\rho\, dv}{r} \tag{12-34}$$

where the integral is taken over the whole of the attracting body and r is the distance from the element dv to the point at which V is the potential.

Problem 43. Find the potential, at points both inside and outside, due to a thin spherical shell of uniform density.

Problem 44. Find the potential due to a sphere whose density is a function of the distance from its center only.

Problem 45. Find the potential, at a point on its axis, due to a thin circular sheet of uniform density.

References

COFFIN, J. G.: "Vector Analysis," John Wiley & Sons, Inc., New York.

GIBBS, J. WILLARD, and E. B. WILSON: "Vector Analysis," Charles Scribner's Sons, New York.

MARGENAU and MURPHY: "The Mathematics of Physics and Chemistry," D. Van Nostrand Company, Inc., New York.

PAGE, L.: "Introduction to Theoretical Physics," Chap. 1, D. Van Nostrand Company, Inc., New York.

CHAPTER XIII

ELECTROSTATICS

The theory of electrostatics is based on a fundamental law of force between point charges. The law is an experimental law, but the range of its direct experimental verification is by no means as great as the range of its application. It is logically possible that the law is only approximate rather than general and exact, but the assumption of exact validity has led to a theory which satisfactorily describes a wide range of phenomena, so that much indirect support of the fundamental law has been obtained.

1. Electrostatic Fields Due to Fixed Charges. *a. The Fundamental Law of Force.*—The law upon which electrostatics is based was verified with considerable accuracy by Coulomb about 1785 and is known as *Coulomb's law.* By means of his torsion balance he found that the force of attraction or repulsion between two small charged bodies is proportional to the product of their charges and inversely proportional to the square of the distance between them. If the charges are of the same kind, or sign, the force is one of repulsion; if the charges are of opposite signs, the force is one of attraction. The existence of the forces and of two kinds of charges had been known for a long time, but Coulomb was the first to establish the quantitative law of force with precision. The accuracy of such a direct experimental verification of the law depends upon the use of test bodies whose linear dimensions are small compared with the distances between them. Later and slightly more indirect experiments have confirmed the law with still greater precision when applied to bodies of ordinary dimensions. The expression for the law of force in a vacuum can be written

$$\mathbf{F} = \frac{1}{4\pi\kappa_0}\frac{q_1 q_2}{R^2}\,\mathbf{e} \tag{13-1}$$

F represents the vector force on one particle due to the other, q_1 and q_2 represent the magnitudes of the electrical charges on the two particles, with the appropriate signs, and R is the distance between them. The unit vector **e** is parallel to the line connecting the two particles and points in such a direction that it represents a repulsion between them when the two charges have the same sign and an attraction when they have opposite signs. The constant κ_0 has dimensions that depend on the dimensions assigned to the electrical charge and a value that depends on the units used in expressing the force, the distance, and the charge. It is written in the denominator and multiplied by 4π for convenience in later equations.

The law of force expressed by equation (13-1) is similar to the law of gravitation but differs from it in two important respects. In the first place, there are two kinds of electrical charge, designated as *positive* and *negative*. Consequently there are forces of both attraction and repulsion, while gravitational forces are only attractive. In the second place, the charges represented by q_1 and q_2 have no connection with the inertia of the particles on which they are located. Only if the particles are electrons or protons themselves are the masses and the charges related.

b. Field Strength and Potential.—An electric field is said to exist at any point at which an electrically charged, stationary body experiences a force because of its charge. In the case of two charges that repel each other, an electric field exists at the location of each charge. The strength of the field at a point, **E**, is a vector. It gives the direction and the magnitude of the force, per unit positive charge, experienced by a small charged particle placed at the point in question. The field is called an *electrostatic field* if its changes with time are so slow that they need not be considered. The utility of this definition of a field depends on the fact that the force on a particle is always proportional to its charge, so that from a specification of the field the force on any charge follows directly.

There are two points of view that may be taken with respect to the nature of an electric field. According to the older

"action-at-a-distance" point of view, the electric field strength is merely a convenient way of representing the forces between all the individual charges. It is then purely a mathematical convenience. On the other hand, according to the point of view developed by Faraday and Maxwell, the electric field represents a state of the "ether" in the region in which the field exists. Since, at the present time, it is generally desired to

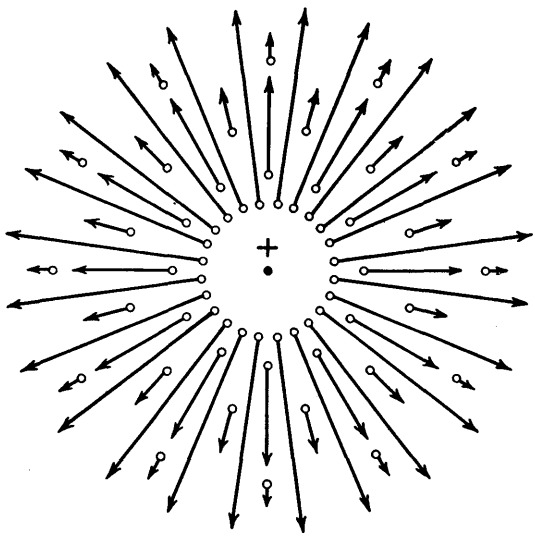


FIG. 13-1.—The electric field due to a point charge is everywhere radial and inversely proportional to the distance from the charge. In this figure the small open circles indicate the points with which the vectors are associated.

avoid the term *ether*, the field is regarded as a certain state of "space." A statement concerning a "state of space" may seem to have little physical content, but an electric field is regarded as having a certain physical reality. In particular, it may possess energy and momentum. When principal attention is paid to the field, the charges that produce it are more or less incidental. This point of view has received much support from the production of electromagnetic waves by means of which energy can be propagated. Since, however, the mathematical development of electrostatics does not depend upon the view adopted as to the nature of the field, the question will be left

open during this chapter and an attempt will be made to keep both possibilities in mind.

In defining the vector **E** as the force per unit charge, it is important to pay attention to the test charge by means of which the force is measured. In the case of a field "produced" by fixed charges, the test charge may be as great as desired if its geometrical extent is small enough. This implies, of course, that there is some way of fixing the charges so they will not be moved by the test charge. To measure a field in the general case, it is necessary to use such a small test charge that its influence in displacing other charges and so distorting the field can be neglected. A single electron is the smallest charge that can be used, and hence it must be concluded that an electric field cannot be measured, and hence cannot be defined, with unlimited precision. Such problems are of importance in the study of atomic electrodynamics and electron theory, but they are of little consequence in the ordinary applications of electricity and will not be considered further in this work.

The law of electrical force between charged particles is enough like Newton's law of gravitation to permit similar use of a potential. A region in an electrostatic field is said to have the potential Φ if

$$\mathbf{E} = -\operatorname{grad} \Phi \tag{13-2}$$

From the law of force it follows that, when the positions of all the charges are known, the potential can be calculated from a formula similar to that used for the gravitational potential

$$\Phi = \frac{1}{4\pi\kappa_0} \sum_i \frac{q_i}{R_i} \tag{13-3}$$

Here q_i is the magnitude of the charge (with the proper sign) that is located at the point i, and R_i is the distance from the point at which Φ is the potential to the charge q_i. The potential is a function of position, so that every point in the field will have associated with it a number which is its value of the potential. Equations (13-1) and (13-2) would still be true if any arbitrary constant were added to the potential as given by (13-3). It is customary, however, to take Φ as given in equation (13-3),

except in some idealized cases where it is convenient to take the potential as zero at some point other than at infinity. Since the charge can be either positive or negative, the potential can have either sign.

c. Electrical Units.—There has been much confusion concerning the units in electrical problems, for numerous different systems have been proposed and used. Each of these has its own particular advantage, but the multiplicity of systems has led to frequent errors in the interpretation of equations. Within recent years still another system has been adopted by numerous international scientific and engineering bodies. It is known as the *Giorgi system* after its proponent or more usually as the mks system after the fundamental units, meter, kilogram, second, that it uses. This system has grown out of the practical electrical units long used by engineers and for this reason is well adapted to engineering computation. Although the system is devised to facilitate computation rather than to promote understanding of the fundamental processes involved, it appears to offer numerous advantages and has been received with a good deal of approval. It will therefore be used as the basis of the material presented here, but reference will also be made to the Gaussian system used in much of the published literature.

If κ_0 is set equal to $1/4\pi$ and is considered to be dimensionless, equation (13-1) defines units of electrical charge in addition to expressing the law of force. When the force is expressed in dynes and the distance in centimeters, the charge will be in electrostatic units whose dimensions follow from equation (13-1). These are the units of charge used in the Gaussian system and in much of the older work on the subject.

In the mks system, the charge is measured in coulombs. It is not necessary to give the definition of a coulomb here but merely to emphasize that it is a unit of charge which can be reproduced by following suitable specifications. It may be regarded as the fourth fundamental unit in a system that might be known as the *meter-kilogram-second-coulomb* (mksc) system. The distance is measured in meters, and the force is measured in newtons. The meter is, of course, a well-known measure of

length, and the newton is the force necessary to give a mass of one kilogram an acceleration of one meter per second per second. Hence one newton is 10^5 dynes.

When the electrical charge is measured in coulombs, the distance in meters, and the force in newtons, the dimensions of κ_0 are given by equation (13-1). The numerical value of κ_0 is given by experiment to be

$$\frac{1}{4\pi\kappa_0} = 8.988 \times 10^9 \frac{\text{newton} \cdot \text{meter}^2}{\text{coulomb}^2} \tag{13-4}$$

It follows that the field **E** is measured in newtons per coulomb and that the potential Φ, introduced in equations (13-2) and (13-3), is measured in newton-meters per coulomb or joules per coulomb. This derived unit is called a *volt*, so that Φ is measured in volts and **E** in volts per meter.

d. Continuous Distribution of Charge.—If there are many charges close together, it is usually convenient to deal with the charge density ρ instead of with the individual point charges. This charge density is equal to the algebraic sum of the charges in a small element of volume dv, divided by dv. Care must be used in the application of this definition. The density of a quantity is usually defined as the limit toward which the above ratio approaches as the size of the volume element is decreased. However, we know that electricity exists in the form of discrete elementary charges, so that the density ρ will not approach a limit as the volume element is reduced to such a size that it contains only two or three charges. The apparent value of ρ determined in this way would depend so much on the exact size and location of dv that it would be of no use.

If dv were reduced still further, so as to include only a portion of a charge, the knowledge of ρ would require a knowledge or assumption about the structure of these ultimate units. In addition, we know through the study of electron theory that these ultimate units of charge, or electrons, are not to be treated as having a precisely defined location and must, in fact, be treated by the methods of quantum mechanics. To avoid such problems and to construct a theory of electricity that is satis-

factory on the macroscopic scale it is sufficient to require that the dimensions of the volume elements used be large compared with interelectronic distances and that such volume elements contain a great many electrons and positively charged nuclei. On the other hand, of course, the dimensions of the volume elements must be small compared with other distances involved. This requirement limits the theory to macroscopic problems. For application to atomic structure, such questions must be reconsidered.

With the above restriction in mind, the density of electricity may be used, and the expression for the potential at a point becomes

$$\Phi = \frac{1}{4\pi\kappa_0} \int \frac{\rho \, dv}{R} \tag{13-3a}$$

The integration is to be taken over all the volume in which there is charge, and in fact over all space, since those parts in which $\rho = 0$ make no contribution to the integral. The validity of equations (13-3) and (13-3*a*) depends upon the fact that the potential due to a group of different charges is the sum of the potentials due to them individually. It is an experimental fact that the force between two charges seems to be entirely independent of the presence or state of motion of other charges, and the forces are to be treated as additive.

The form of equation (13-3*a*) shows that, as long as the charge density ρ does not become infinite, the potential Φ is a continuous function of position and can be differentiated to give the field.

Problem 1. Two equal positive charges are placed at opposite corners of a square, and two negative charges of the same magnitude are placed at the other two corners. Find the potential and the field strength at points near the center of the square. Include points outside the plane.

Problem 2. Three charges are located at the vertices of an equilateral triangle. Show that the force exerted by any two upon the other one is directed through the center of charge. The definition of the center of charge is analogous to that of the center of mass.

Problem 3. Find the potential and the electrostatic field around a uniformly charged, infinite, straight wire. This problem can be

treated by first considering the wire to be finite, finding the potential about its center, and then letting its length increase.

Problem 4. Find the potential and the field due to two concentric spherical shells that are differently charged.

e. Restricted Form of Gauss's Law.—Gauss's law states that the integral of the electric field over any closed surface is equal to the total charge inside the surface divided by κ_0. This is called the *restricted* form of Gauss's law because it is valid only

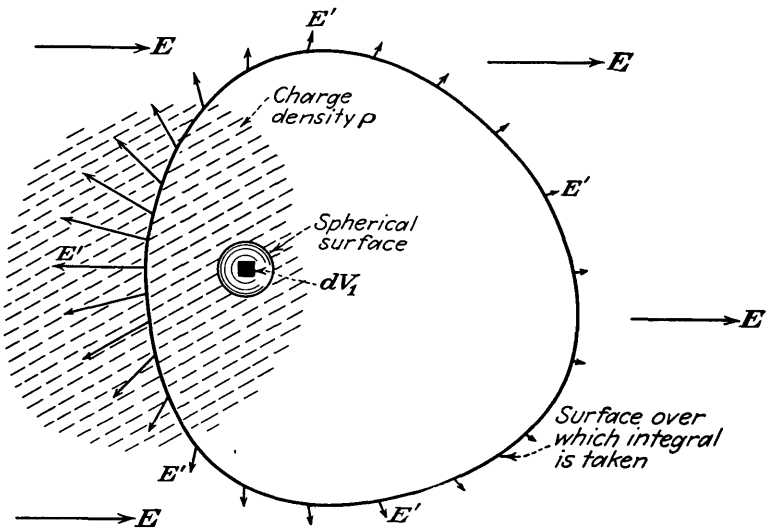


FIG. 13-2.—The field E' is that due only to the charge in the volume element dv_1.

in the absence of polarization, which will be discussed later. In symbols,

$$\int_S \mathbf{E} \cdot \mathbf{ds} = \frac{q}{\kappa_0} = \frac{1}{\kappa_0} \int_V \rho \, dv \qquad (13\text{-}5)$$

To show this, the field at each point of the surface over which the integration is to be carried out may first be divided into two parts. One part is that due to the charge in a certain volume element dv_1 and may be called $\mathbf{E}_1$. The other part is due to all the rest of the charge and may be called $\mathbf{E}$. The law is first to be established for $\mathbf{E}_1$ only.

A spherical surface is described about the volume element dv_1, and the field on this surface due to $\rho \, dv_1$ may be easily calcu-

lated. The integral over this surface can be shown by actual integration to be equal to $\rho\, dv_1/\kappa_0$. By Gauss's theorem and the fact that the divergence of $\mathbf{E}_1$ is zero outside of dv_1, it follows that the integral of $\mathbf{E}_1$ over S is also $\rho\, dv_1/\kappa_0$ if dv_1 and the spherical surface are inside S but is zero if dv_1 and spherical surface are outside S. Hence the law is established for $\mathbf{E}_1$ and the charge $\rho\, dv_1$ that may be said to produce $\mathbf{E}_1$.

Since the total field is the sum of the fields due to individual elements of charge, the law is established for the whole field.

It is sometimes possible to use Gauss's law and the symmetry of a problem to evaluate the field strength itself. Consider the case of an infinite straight line of charge with no other charges present. Due to the symmetry of the problem the electric field must be everywhere radial, and its magnitude can depend only on the distance from the wire. Hence $\mathbf{E} = \mathbf{e}_r E(r)$, where $\mathbf{e}_r$ is a unit vector in the radial direction.

Construct a circularly cylindrical surface with the line of the charge as its axis and with plane ends perpendicular to the axis. Let the length of the cylinder be dl and the radius r. The electric field will be parallel to the ends of the cylinder so that on those surfaces $\mathbf{E} \cdot \mathbf{ds} = 0$, and it will be normal to the cylindrical surface. Hence the total surface integral is $2\pi r E\, dl$. By Gauss's law, this is equal to the total charge inside, or $\lambda\, dl/\kappa_0$, where λ is the linear density along the line. Hence

$$E = \lambda/2\pi\kappa_0 r$$

Problem 5. Use Gauss's law to find the field due to an infinite plane uniform distribution of charge.

Problem 6. Use Gauss's law to show that the field is zero inside a spherical shell of uniform surface charge density.

From Gauss's theorem as discussed in the previous chapter, equation (13-5) can be written

$$\int_S \mathbf{E} \cdot \mathbf{ds} = \int_V div\, \mathbf{E}\, dv = -\int_V \nabla^2\Phi\, dv = \frac{1}{\kappa_0}\int \rho\, dv \qquad (13\text{-}6)$$

Since this is true for any volume, however small, it is true for each individual volume element and there results one of the fundamental laws of electricity,

$$-\nabla^2\Phi = div\, \mathbf{E} = \rho/\kappa_0 \qquad (13\text{-}7)$$

f. Dipoles.—In addition to the electric fields due to isolated charges and to volume distributions of charge, electrical phenomena are observed that can best be described in terms of *electric dipoles.* In terms of the microscopic theory of matter and electricity, the necessity for considering these distributions is due to the limitations mentioned above on the concept of a charge distribution. In a macroscopic theory, however, it is not necessary to go into the origin of such distributions but merely to take them as observed.

A dipole may be visualized as a limiting case of two equal and opposite charges separated by a distance d. Let R be the distance from the charges to a point at which the field is to be computed. When the ratio d/R approaches zero, the two charges can be regarded as a dipole. Such a dipole is described by its vector moment $\mathbf{p}$. This moment has a magnitude given by the product of one of the charges and the distance between them. Its direction is such as to point from the negative to the positive charge.

The potential due to a dipole is given by

$$\Phi = \frac{1}{4\pi\kappa_0}\frac{\mathbf{p}\cdot\mathbf{R}}{R^3} \tag{13-8}$$

where $\mathbf{R}$ is the vector from the center of the dipole to the point at which Φ is the potential. It must always be remembered that a dipole is a limiting case. Two equal and opposite charges constitute a dipole only when all observations are made so far away that the separation between the charges can be neglected.

Problem 7. Show that equation (13-8) gives the potential due to a dipole.

Problem 8. Show that the field strength due to a dipole is given by

$$\mathbf{E} = -\frac{1}{4\pi\kappa_0}\frac{\mathbf{p}}{R^3} + \frac{3}{4\pi\kappa_0}\left(\frac{\mathbf{p}\cdot\mathbf{R}}{R^5}\right)\mathbf{R} \tag{13-9}$$

Problem 9. Show that, if a dipole is parallel to the polar axis, the components of the electric field parallel and perpendicular to the radius vector are

$$E_r = \frac{1}{4\pi\kappa_0}\frac{2p}{R^3}\cos\theta \qquad \text{and} \qquad E_\theta = \frac{1}{4\pi\kappa_0}\frac{p}{R^3}\sin\theta \tag{13-10}$$

Problem 10. Show that the field of a dipole has components parallel and perpendicular to the dipole equal to

$$E_{\parallel} = \frac{1}{4\pi\kappa_0}\frac{p}{R^3}(3\cos^2\theta - 1) \qquad \text{and} \qquad E_{\perp} = \frac{1}{4\pi\kappa_0}\frac{3p}{2R^3}\sin 2\theta \quad (13\text{-}11)$$

where θ is the angle between the axis of the dipole and the vector from the dipole to the point at which the field is measured.

g. Electric Polarization.—If an element of volume is filled with dipoles, its net charge will be zero and yet it may give rise to an electric field. In this case equation (13-3*a*) is inadequate

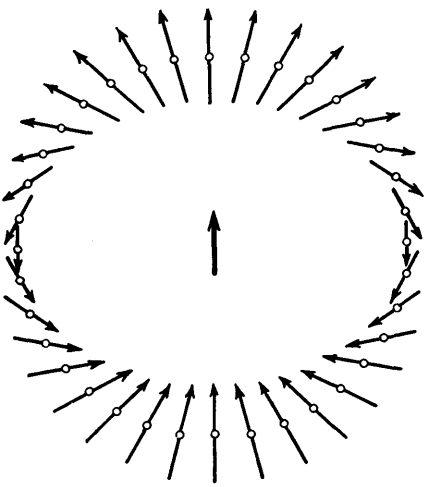


FIG. 13-3.—The field due to a dipole is inversely proportional to the third power of the distance from the dipole. This figure shows the field vectors at one distance.

for computing the potential, and additional account must be taken of the dipoles. The fact that this case exists is associated with the limitations on the volume elements that can be used. They are not infinitesimal in the mathematical sense but have a finite size. They cannot be made small enough to include only one of the charges of a dipole because of the limitations imposed by quantum mechanics. This limitation has already been mentioned in connection with the discussion of charge density, and it sometimes leads to situations in which the charge density is zero but in which there is an electric field that can be described as due to a distribution of dipoles.

This deficiency in equation (13-3*a*) can be remedied by

adding to it an expression for the potential due to the dipoles themselves. If $\mathbf{P}$ represents the resultant dipole moment per unit of volume, this additional potential is given by

$$\Phi_P = -\frac{1}{4\pi\kappa_0}\int \frac{\mathbf{P}\cdot\mathbf{R}}{R^3}\,dv \qquad (13\text{-}12)$$

where $\mathbf{R}$ is the vector from the point at which Φ_P is the potential to dv. This is opposite to the direction used in equation (13-8) and is the reason for the negative sign. The vector $\mathbf{P}$ is called the *electric polarization.* It is a vector function of position such that $\mathbf{P}\,dv$ represents the vector sum of the dipoles in the element of volume dv. As is the case with the charge density ρ, the polarization $\mathbf{P}$ is defined subject to the limitation that the volume elements must always contain a large number of atoms.

The integral in equation (13-12) is taken over all the volume in which a polarization exists, or over all space, since the regions in which $\mathbf{P} = 0$ make no contribution to the integral.

Problem 11. Consider a long thin rod of constant cross section S, uniformly polarized in the direction of its axis. Use equation (13-12) to find the electric field due to this distribution of polarization.

Problem 12. Consider a right circular cylinder of radius 10 cm and length 15 cm filled with a uniform polarization $\mathbf{P}$ parallel to the axis. Find the electric potential at points on the axis both inside and outside of the cylinder by the use of equation (13-12).

If the polarization $\mathbf{P}$ is everywhere finite with finite derivatives, it is possible to transform equation (13-12) into a form that may be easier to use. By considering $div\,(\mathbf{P}/R)$ and equating its volume integral to a surface integral by Gauss's theorem, it follows that

$$\Phi_P = \frac{-1}{4\pi\kappa_0}\int \frac{\mathbf{P}\cdot\mathbf{R}}{R^3}\,dv$$

$$= \frac{-1}{4\pi\kappa_0}\int \frac{div\,\mathbf{P}}{R}\,dv + \frac{1}{4\pi\kappa_0}\int \frac{\mathbf{P}\cdot\mathbf{ds}}{R} \qquad (13\text{-}13)$$

where the volume integral is taken over any volume that includes all the polarization and the surface integral is taken over the surface enclosing this volume. Such a surface can be selected so that $\mathbf{P} = 0$ on it and hence so that the surface integral

vanishes. It follows then that

$$\Phi_{\mathbf{P}} = \frac{-1}{4\pi\kappa_0} \int \frac{div\, \mathbf{P}}{R}\, dv \qquad (13\text{-}13a)$$

The form of (13-13a) compared with that of (13-3a) shows that the negative divergence of the polarization has the same effect in producing an electric field as does an electric charge density (see Fig. 13-4).

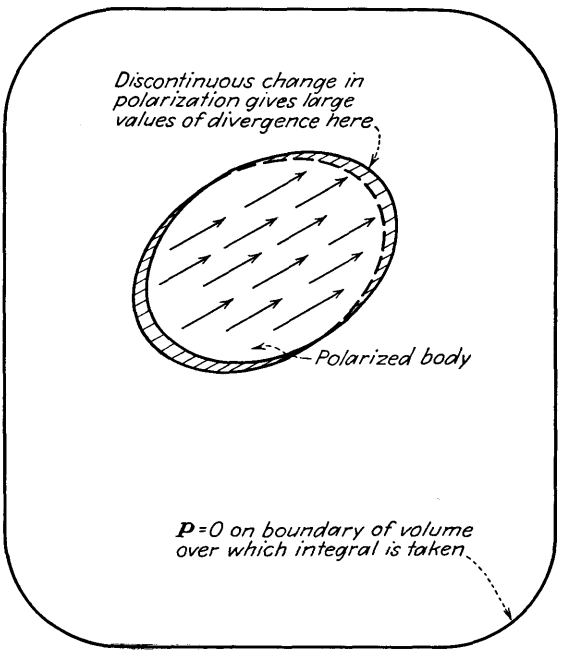

Discontinuous change in polarization gives large values of divergence here
Polarized body
P=0 on boundary of volume over which integral is taken

FIG. 13-4.—The very rapid rate of change of the normal component of polarization at the surface of a polarized body has the same effect in producing a field as has a surface charge.

In Probs. 11 and 12 the polarization has a constant value, and hence a zero divergence, within the regions where it differs from zero. The application of equation (13-13a) to such a region would then seem to lead to a zero potential. This is clearly incorrect. The difficulty lies in the fact that the divergence of the polarization is not everywhere finite as is assumed in in the derivation of equation (13-13a).

The apparent discontinuity in the polarization at the surface of a polarized body is probably an idealization. In any case

it is possible to retain all the significant features by treating this change as very rapid, but not discontinuous. Equation (13-13*a*) is then applicable, but care must be taken to include this very large value of the divergence near the surface. The volume integral of this part of the divergence can then be approximated by a surface integral.

Let the x and y axes lie in the surface at a selected point. The integrand in this neighborhood will then be

$$-\frac{1}{R}\left(\frac{\partial P_x}{\partial x}+\frac{\partial P_y}{\partial y}+\frac{\partial P_z}{\partial z}\right)dx\,dy\,dz$$

$$=-\frac{1}{R}(P_{z1}-P_{z2})dx\,dy \quad (13\text{-}13b)$$

The derivatives parallel to the surface are considered as negligible compared with the derivative normal to the surface. The change in R is also negligible. P_{z1} is the normal component of the polarization on one side of the surface, and P_{z2} is the normal component on the other side. Hence, if the distribution of polarization contains surfaces of discontinuity, the integral of the divergence over these surfaces can be replaced by a surface integral of the normal component of the discontinuous change in potential.

It is very easy to confuse the surface integral on the right-hand side of equation (13-13*b*) with the surface integral neglected in equation (13-13). With the proper understanding, and the proper selection of volumes and surfaces, they lead to the same result. It seems, however, that a more universally applicable procedure for handling problems of polarization can be reached by (1) regarding surfaces of discontinuity as idealizations that should properly be replaced by a very rapid change; and (2) replacing the volume integral over this region of very rapid change by a surface integral of the normal component of the total discontinuity.

Problem 13. Show how the distributions of polarization in Probs. 11 and 12 can be represented by distributions of apparent surface charge.

By combining equations (13-3*a*) and (13-13*a*) it can be

seen that the total potential, due to both charge and polarization, is given by

$$\Phi = \frac{1}{4\pi\kappa_0} \int \frac{\rho - div\,\mathbf{P}}{R}\, dv \tag{13-14}$$

When the distribution of charge and of polarization is completely known, the potential in any electrostatic situation can be evaluated by the use of equation (13-14). In the use of this equation it must be remembered that the integral is to be carried out over all space and must include all of the charge and polarization that affects the potential. In some cases there will be discontinuities in the polarization, but it will be assumed that these can be approximated by a rapidly changing, continuous distribution, as already described.

The fact that the divergence of the polarization and the normal component of the polarization on a surface of discontinuity are equivalent to volume and surface charge densities has led to numerous attempts to visualize the effect in terms of the charges composing the dipoles. If all dipoles are parallel to each other and normal to the surface, one may picture the positive or the negative ends of these dipoles as being the positive or negative surface charges. Such visualization clearly violates the restrictions imposed on the size of the volume elements that can be used and must be employed with caution.

In a similar fashion the ρ in equation (13-14) is sometimes referred to as a "real" charge density while the $-div\,\mathbf{P}$ is called a "bound" charge density. This also seems of doubtful value, and it appears more practicable to recognize that the integral is just a method of calculating the potential due to both charges and dipoles.

h. Complete Form of Gauss's Law.—In equation (13-14) the quantity $\rho - div\,\mathbf{P}$ takes the place of ρ in equation (13-3*a*). By the same type of reasoning as led to equation (13-5) it can be shown that

$$\int \mathbf{E} \cdot \mathbf{ds} = \frac{1}{\kappa_0} \int (\rho - div\,\mathbf{P}) dv$$

$$= \frac{1}{\kappa_0} \int \rho\, dv - \frac{1}{\kappa_0} \int \mathbf{P} \cdot \mathbf{ds} \tag{13-15}$$

From this it follows that

$$\int(\kappa_0\mathbf{E} + \mathbf{P}) \cdot \mathbf{ds} = \int\mathbf{D} \cdot \mathbf{ds} = \int\rho\, dv \qquad (13\text{-}16)$$

This is the complete form of Gauss's law that must be used when polarization is present. The vector $\mathbf{D} = \kappa_0\mathbf{E} + \mathbf{P}$ is called the *electric displacement.* It is of importance largely because of the role it plays in Gauss's law.

It follows directly from equation (13-16) that

$$div\ \mathbf{D} = \rho \qquad (13\text{-}17)$$

and this may be called the *differential* form of Gauss's law.

i. Summary of Electric Fields Due to Known Distributions of Charge and Polarization.—The discussion thus far has referred to the case in which the electric charges and electric dipoles are fixed in known positions. The question as to whether they are held in these positions by nonelectrical forces or are in equilibrium under the action of electrical forces only has not been considered. If the charge distribution is known, ρ is known as a function of position and any isolated point charges are in known positions. The knowledge of the distribution of polarization requires the knowledge of $\mathbf{P}$ as a function of position and of the locations of any isolated dipoles. When these are given, the potential at every point in space may be defined as

$$\Phi = \frac{1}{4\pi\kappa_0}\int\frac{1}{R}(\rho - div\ \mathbf{P})dv \qquad (13\text{-}14)$$

subject to the restrictions already discussed as to the minimum size of the elements dv. This integral will give the potential Φ a value that is everywhere continuous and single-valued when ρ and $div\ \mathbf{P}$ are always finite. It applies for points inside as well as outside of the distributions of charge and polarization.

When the potential Φ is known, the electric field can be evaluated by the operation of taking the gradient.

$$\mathbf{E} = -\ \mathbf{grad}\ \Phi \qquad (13\text{-}2)$$

This may be taken as the definition of the electric field $\mathbf{E}$. In regions of space accessible to a test charge, this field may be measured by the force on such a charge. Inside of dielectrics,

however, **E** may be regarded as that function of the charge and polarization distributions given by equations (13-14) and (13-2).

The displacement **D** is defined by

$$\mathbf{D} = \kappa_0 \mathbf{E} + \mathbf{P} \tag{13-18}$$

It follows from this definition that the electric displacement in a vacuum is κ_0 times the electric field, and therefore the field and the displacement are assigned different dimensions. This is,

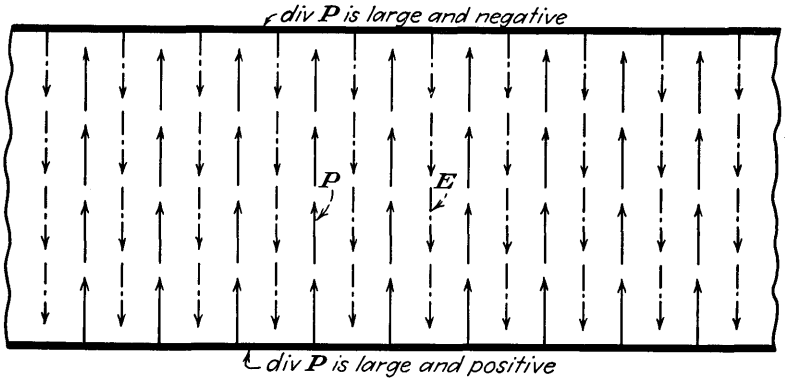


FIG. 13-5.—Polarization and field inside a uniformly polarized slab.

of course, a logical possibility, but it may also be misleading and is probably the only significant objection to the mksc system of units as now used. In the older Gaussian units, $\mathbf{D} = \mathbf{E} + 4\pi\mathbf{P}$, and it is much more obvious that the displacement is obtained from the field merely by adding the polarization. Nevertheless, equation (13-18) should cause no trouble if the apparent role of κ_0 as the "permittivity of a vacuum" is not taken as suggesting the possibility of polarization in a vacuum.

Problem 14. Consider a thick slab of uniform polarization, perpendicular to the plane surfaces. Let the area of the slab be indefinitely large. Evaluate the potential, the field, and displacement inside and outside of the slab. This can be treated by considering a cylindrical volume with its axis parallel to the direction of polarization. One flat end is inside the slab, and the other outside. From a consideration of the symmetry of the situation and the complete form of Gauss's law, the problem can be solved (Fig. 13-5).

Problem 15. Consider a charge Q at the center of a thick spherical shell of polarization. Let the polarization be everywhere along the radius and constant in magnitude. Find the potential, field, and the displacement at all points, and illustrate Gauss's law by integrating over spherical surfaces inside and outside of the shell.

2. The Effect of an Electrostatic Field on Material Bodies.— All matter is now believed to be made up of electric charges in the form of electrons and atomic nuclei. If a material body is placed in an electric field, it may be expected that the field will cause a rearrangement of the charged particles and, in general, that the body as a whole will experience a force. The complete description of the rearrangement cannot be given in terms of the ordinary laws of electricity. Recourse must be had to the methods of quantum mechanics. Nevertheless, the macroscopic aspects of the response of material bodies to electric fields can be satisfactorily described by dividing the bodies into two classes, known, respectively, as *conductors* and as *insulators.*

The distinction between conductors and insulators is not sharp, and some substances might be put in either class. Nevertheless, it is a very useful distinction. It emphasizes the two ways in which substances react to an electric field.

If a conductor is placed in an electric field, some of the electrons in it will move through the body until they build up surface charges that will completely neutralize the field originally existing inside the conductor. On this account, there is, by definition, no electric field inside a conductor. This statement refers, of course, only to the static situation after equilibrium has been attained. There does exist a field inside the conductor while the electrons are moving around to neutralize it. In addition, a conductor contains no dipoles and hence no polarization. A conductor may then be defined as a substance inside of which there exists no electric field and no polarization, in static situations.

An insulator, on the other hand, is a substance in which all the electrons are held at or near to positions of equilibrium and in which the effect of an electric field is to displace them slightly and hence to produce a polarization. An insulator is a sub-

stance in which the polarization at each point is some function of the electric field at the same point. In some cases the polarization also depends on the past history of the electric field.

a. Conductors.—The essential property of a conductor is given by the definition above. In an electrostatic situation, all parts of a conductor will be at the same potential, since the field inside it is zero. The potential everywhere is given by equation (13-14). If there are outside distributions of charge and polarization that would tend to cause different parts of the conductor to be at different potentials, the electrons in the conductor will move until the potential produced by them combines with the externally produced potential to give zero field and a constant potential throughout the conductor. In general, this requires a nonuniform distribution of charge over the surface, but in a few simple cases the symmetry of the situation is such that a uniform surface distribution of charge produces the necessary result.

In dealing with conductors the charge is usually treated as a surface charge. This is not strictly true; but since the charge is distributed in a layer that is as thin as or thinner than the minimum allowable element of length, all the essential properties are retained when it is treated as a surface charge.

Problem 16. Show from the definition of a conductor and the fact that the potential as given by equation (13-14) is a continuous function of position that the tangential component of the electric field just outside the surface of a conductor is zero.

Problem 17. Show from Gauss's law and the definition of a conductor that any charge on a conductor is at its surface.

Problem 18. Show from Gauss's law that the normal component of the displacement in a dielectric adjacent to the surface of a conductor is given by

$$D_n = \sigma \tag{13-19}$$

where σ is the surface density of the charge.

Problem 19. Consider a thick spherical shell of a conductor on which there is no net charge. If a charge Q is placed at the center, make use of the spherical symmetry of the problem and the fact that the field inside the conductor must be zero to compute the charge dis-

tribution on the conductor. Then compute the field due to the induced charge alone.

As can be seen in Prob. 19 the potential of a conductor is determined not only by any charge that may be on it but also by any other charges or polarization that may be in the neighborhood. If a conductor is completely isolated, its potential is determined by its own charge only. The ratio between the net charge on such a conductor in coulombs and the potential of the conductor in volts is known as its *capacitance.* The capacitance of an isolated sphere is equal to $4\pi\kappa_0 R$, where R is the radius in meters and the capacitance is measured in farads. A farad is one coulomb per volt and hence κ_0 can be expressed in farads per meter.

The term *capacitance* is used in a number of slightly different senses, but in each case it refers to a charge that will produce a unit potential. The above-defined capacitance of an isolated sphere could also be regarded as the capacitance with reference to an enclosing conductor at infinity. For most practical purposes the earth serves as the conductor at infinity, and thus such a capacitance is sometimes referred to as the *capacitance to ground.*

If two conductors are isolated from other bodies or are enclosed in an infinitely large conducting shell, a mutual capacitance may be defined. This is the charge on one conductor divided by the difference in their potentials, when the two conductors have equal and opposite charges. Two conductors so close together that their difference in potential is little affected by other bodies constitute a capacitor, or a condenser, whose capacitance is this mutual capacitance.

Problem 20. Show that the capacitance of an isolated sphere is equal to $4\pi\kappa_0 R$ farads when R is its radius in meters.

Problem 21. If the insulation of the air breaks down at 2×10^6 volts per meter, what is the maximum potential to which an isolated sphere surrounded by air can be charged?

Problem 22. A conducting sphere is placed inside and concentric with a spherical shell that is connected to earth. Show that, if the inner sphere carries a charge q, the outer will carry a charge $-q$. Find the capacitance of this system.

Problem 23. In a capacitor composed of two flat parallel conductors separated by a distance that is small compared with their linear dimensions, the charge will be almost uniformly distributed over the opposing conducting surfaces. Under the assumption that the charge is uniformly distributed and that there are equal and opposite charges on the two plates, show that the capacitance of the condenser is

$$C = \frac{\kappa_0 A}{d} \tag{13-20}$$

where A is the area of one of the plates and d is the distance between them.

b. Insulators or Dielectrics.—If an uncharged insulator is placed in a field due to fixed charges, polarization will be produced throughout its volume. The total electric field at any point will then be different from the field due to the fixed charges. In addition to the original field there will be the field due to the polarization "induced" throughout the dielectric. Hence the field at any point that is effective in producing polarization is determined, in part, by the polarization it produces.

In numerous simple dielectric substances, the polarization is approximately proportional to the electric field, and a dimensionless constant of proportionality χ_e, called the *dielectric susceptibility*, is defined by $\mathbf{P} = \chi_e \kappa_0 \mathbf{E}$. In these substances the displacement also is proportional to the electric field so that

$$\mathbf{D} = \kappa_0 \mathbf{E} + \chi_e \kappa_0 \mathbf{E} = \kappa \mathbf{E} \tag{13-21}$$

and

$$\kappa = K\kappa_0 = (1 + \chi_0)\kappa_0$$

The quantity κ has the same dimensions as κ_0 and is called the *permittivity* of the medium, while K is a dimensionless ratio called the *dielectric constant.*

In the simplest cases of anisotropic crystals in weak fields, the displacement is a linear vector function of the electric field. Thus its direction may be different from that of the field itself. In more complicated cases the polarization and the displacement may be still more complicated nonlinear functions of the

electric field. The relationship between the displacement and the field at any point in a dielectric is a characteristic property of the substance and can be found empirically by properly designed experiments. Only in the very simplest isotropic cases is the dielectric constant as defined in equation (13-21) a useful quantity.

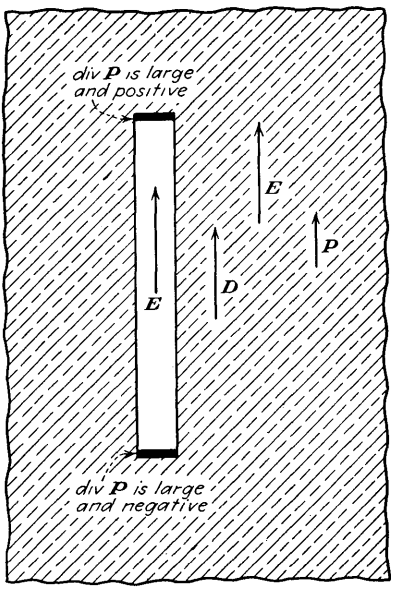


FIG. 13-6.—The field inside a long narrow cavity, whose axis is parallel to the polarization, is equal to the field in the body itself.

Problem 24. Consider a simple isotropic dielectric in which the displacement is parallel to the field. Let a small cylindrical cavity be cut in the dielectric, with its axis parallel to the field and a diameter very much smaller than its length. Let the whole cavity be so small it has no influence on the distribution of field and polarization in the body, and show that the field at the center of the cavity will be equal to the field in the material just outside the cavity (Fig. 13-6).

Problem 25. Under the conditions of the above problem except that the length of the cavity is very much less than the diameter, show that the field at the center of the cavity is equal to $1/\kappa_0$ times the displacement adjacent to it.

Problem 26. Show that the electric field at the center of a small spherical cavity in a dielectric is equal to $\mathbf{E} + \mathbf{P}/3\kappa_0$, where $\mathbf{E}$ and $\mathbf{P}$ are the field and polarization in the dielectric adjacent to the hole.

Problem 27. Show by means of Gauss's law that, at a boundary between two dielectrics on which there is no charge, the normal component of the electric displacement is continuous across the surface.

Problem 28. Show from the fact that the electric field is the gradient of a potential that the tangential component of **E** is continuous across the boundary between any two bodies.

Problem 29. For cases in which equation (13-21) holds, use the results of the above two problems to establish a law of refraction for the lines of electrical force at a boundary between two dielectrics.

Problem 30. Show that at the surface between two dielectrics the electric field behaves as though it were in a vacuum and there were a surface charge of density

$$\sigma = \kappa_0 E_1 \left(\frac{K_1}{K_2} - 1 \right)$$

K_1 and K_2 are the respective dielectric constants of the two media, and E_1 is the normal component of the field in the first medium. The positive direction of E_1 is taken from the first medium to the second.

Problem 31. A flat-plate condenser is attached to a battery so that the two plates are kept at a constant difference of potential. At first the plates are in a vacuum. Compute the surface density of charge on the plates and the field between them. A dielectric of constant K is then inserted between the plates so that it practically fills the whole space but does not quite touch either plate. Find the distribution of charge, polarization, field, and potential in this case. The effects of the edges should be neglected in all cases, and, owing to the symmetry of the problem, the field and the polarization can be assumed perpendicular to the plates.

Problem 32. Carry through the above problem when the connection with the battery is broken before the dielectric is inserted.

Problem 33. Show that the capacitance of a flat-plate condenser is multiplied by K when a dielectric described by equation (13-21) is placed between the plates.

3. The General Electrostatic Distribution Problem.—In the first section of this chapter the problems treated were those in which distributions of electric charge and polarization were given and the problem was to compute the resulting potentials and electric fields. In the second section the behavior of conductors and dielectrics under the influence of an electric field was described. The behavior consists in the redistribution of

the charge on the conductors and the production of polarization in the dielectrics. When a conductor or a dielectric is placed in an electric field, the redistributed charge on the conductor, or the polarization produced in the dielectric, leads to additional terms in the expression for the potential and to possible changes in the original field. These again react on the material bodies, and the final result is an equilibrium situation whose determination is the object of the analysis of an electrostatic problem. In general, the distribution of charge and polarization is not given, and the properties of the insulators and dielectrics must be used to find it. The general problem of electrostatics is then to find this distribution of charge and polarization, of potential and field. The minimum information necessary for this purpose depends upon the situation.

a. The fundamental equation of electrostatics is the differential form of Gauss's law, equation (13-17),

$$div\,\mathbf{D} = \rho \tag{13-17}$$

In a homogeneous dielectric in which the displacement is proportional to the field, with a permittivity κ,

$$div\,\mathbf{E} = \frac{\rho}{\kappa} \tag{13-17a}$$

In the discussion that follows only this idealized type of dielectric will be considered.

In most problems one has to deal with discrete pieces of homogeneous dielectric to which equation (13-17*a*) can be applied and it is not necessary to use the more general form of (13-17). Since the field strength is the negative gradient of the potential, equation (13-17*a*) is equivalent to Poisson's equation

$$\nabla^2\Phi = \frac{-\rho}{\kappa} \tag{13-17b}$$

In most problems the charge is confined to isolated points or to the surfaces of conductors. An exception is the case of a space charge in an electron tube, but such cases require special methods of treatment and will not be considered here. The

charge distribution in the usual case can be included in the boundary conditions, and thus the potential will be a continuous function that satisfies Laplace's equation

$$\nabla^2\Phi = 0 \tag{13-22}$$

in each dielectric. The connection between the potential in one dielectric and in the next is made by the appropriate schedule of boundary conditions.

In the remainder of this chapter it will be assumed that

1. All charges are located on conductors or at specified isolated points.

2. The material bodies consist of discrete pieces of conductor or of homogeneous dielectric with a permittivity κ.

3. The only polarization is that induced in the dielectrics.

b. Schedule of Boundary Conditions.—In each of the discrete regions of space defined by the conditions of the problem, the potential satisfies Laplace's equation. Its behavior at the boundaries and at isolated points is given by the following conditions:

1. The potential is continuous across the boundaries.

2. *a.* The normal component of the electric displacement is continuous across the boundary between two dielectrics, when there is no charge on the boundary.

$$\kappa_1 \frac{\partial\Phi_1}{\partial n} = \kappa_2 \frac{\partial\Phi_2}{\partial n} \tag{13-23}$$

where n represents a normal to the surface.

b. At the boundary between a conductor and a dielectric

$$\kappa \frac{\partial\Phi}{\partial n} = -\sigma \tag{13-24}$$

where σ is the surface density of charge on the conductor and n is the unit normal pointing into the dielectric.

3. On the surface of a conductor one or the other of the following is specified:

a. The potential of the conductor.

b. The total charge on the conductor.

4. *a.* In the neighborhood of an isolated charge the potential has the form of $(Q/4\pi\kappa R)$ + constant.

b. In the neighborhood of an isolated dipole the potential has the form of $(\mathbf{p} \cdot \mathbf{R}/4\pi\kappa R^3)$ + constant.

5. At large distances from all charges and polarization the potential vanishes at least as fast as $1/R$. It is sometimes desired, however, to treat a portion of a problem, in which case this condition may be replaced by some other. For example, if a number of bodies are placed between the flat plates of a very large capacitor, the field at infinity may be said to be a uniform field.

c. Theorem of Uniqueness.—The solution of problems in electrostatics is very much simplified by the theorem of uniqueness. According to this, if any function Φ satisfies Laplace's equation and the above schedule of boundary conditions, it is unique. No other function will satisfy the conditions, with the possible exception of those which differ immaterially from the first by an additive constant.

To prove this theorem, let Φ_1 and Φ_2 be two possible potentials for an electrostatic situation, and let $U = \Phi_1 - \Phi_2$. The theorem is proved when it can be shown that U must be either zero or a constant.

It is first clear that the theorem is satisfied throughout each conductor. If the potential of the conductor is specified, $U = 0$. If it is not specified, the definition of a conductor still requires U to be constant. It is also clear that $\nabla^2 U = 0$ throughout each of the dielectrics, since both Φ_1 and Φ_2 must satisfy Laplace's equation. Furthermore, it follows from the conditions on the potential at infinity that U either vanishes or is a constant. It then remains to investigate the behavior of U throughout the dielectrics.

By applying Gauss's theorem to the vector $(U \operatorname{\mathbf{grad}} U)$, it can be shown that within each dielectric

$$\int_V U\nabla^2 U \, dv - \int_S U \operatorname{\mathbf{grad}} U \cdot \mathbf{ds} + \int_V (\operatorname{\mathbf{grad}} U)^2 \, dv = 0 \quad (13\text{-}25)$$

The first term in this equation is zero because $\nabla^2 U = 0$, so that

$$\int_V (\mathbf{grad}\ U)^2\, dv = \int_S U\ \mathbf{grad}\ U \cdot \mathbf{dS} \qquad (13\text{-}26)$$

This equation can be multiplied by κ and the corresponding equations for all the different dielectrics added together. Then

$$\sum_i \kappa_i \int_{V_i} (\mathbf{grad}\ U)^2\, dv = \sum_i \kappa_i \int_{S_i} U\ \mathbf{grad}\ U \cdot \mathbf{dS} \qquad (13\text{-}27)$$

The surface integral is taken over the surface of each portion of dielectric with **dS** in the direction of the outward-drawn normal.

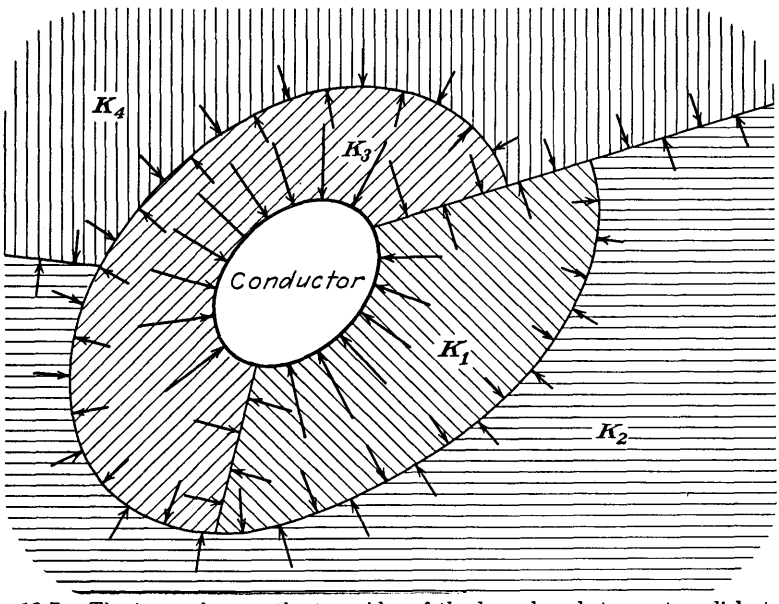


FIG. 13-7.—The integrals over the two sides of the boundary between two dielectrics cancel each other.

U and **grad** U are the values of the functions just inside the surface. The various surfaces will have the following types of properties (Fig. 13-7):

1. Surfaces at infinity. On these either U or **grad** U becomes so small that the surface integral vanishes.

2. Surfaces where two dielectrics meet. On these the integral on one side will be the negative of the integral on the other, since U must be continuous across the boundary, and

the normal component of κ **grad** U is also continuous across the boundary.

3. Surfaces where the dielectric meets a conductor. If the potential of the conductor is specified, $U = 0$ on this boundary. If the potential is not specified, the integral over all the surfaces adjacent to this conductor must be considered. U is a constant because of the conductor, and $\sum \kappa_i U \int$ **grad** $U \cdot$ **dS** will vanish because of condition 3. Hence the right side of (13-27) will vanish; and, since the integrand on the left can never be negative, **grad** U must everywhere be zero. This requires that U be zero or a constant.

Problem 34. Derive equation (13-25) for any continuous scalar function U.

Problem 35. Consider a spherical conductor carrying a charge Q and surrounded by a dielectric spherical shell of thickness T. Find the potential, field, and displacement at all points.

Problem 36. Show that, in a space containing no charges and surrounded by a conductor, the potential is a constant equal to the potential of the conductor.

Problem 37. Consider a charged spherical conductor of radius A. Outside of the sphere, on one side of a plane through its center is a dielectric of constant K_1 and on the other side a different dielectric of constant K_2. Find the potential, field, and displacement in the dielectric; and find the distribution of the charge over the sphere. Then evaluate the potential by means of equation (13-14).

d. Spherical Harmonics.—A function that satisfies Laplace's equation is called a *harmonic function.* There are many kinds of such functions, and it is important to know those whose symmetry makes their application relatively simple. One group of such functions consists of the spherical harmonics and is especially useful in solving problems with spherical symmetry.

In spherical polar coordinates, Laplace's equation becomes

$$\frac{\partial}{\partial r}\left(r^2 \frac{\partial \Phi}{\partial r}\right) + \frac{1}{\sin \theta} \frac{\partial}{\partial \theta}\left(\sin \theta \frac{\partial \Phi}{\partial \theta}\right) + \frac{1}{\sin^2 \theta} \frac{\partial^2 \Phi}{\partial \varphi^2} = 0 \quad (13\text{-}28)$$

A solution of this equation, whose dependence on the radius is expressed by a single power of r, is a spherical harmonic. There

are a great many different solutions of equation (13-28). Since the equation is linear, the general solution is a linear combination of all the independent particular solutions. We shall treat here only the very simplest cases of elementary problems that can be solved by means of spherical harmonics, for a complete discussion of the properties of these functions constitutes an extensive branch of mathematics.

To find a particular solution of equation (13-28) assume it to be a product of functions of the three independent variables r, θ, φ. This gives the form

$$\Phi = R(r)\Theta(\theta)\phi(\varphi) \tag{13-29}$$

Substitution of this form into (13-28) shows that it is a solution provided that

$$\frac{d^2R}{dr^2} + \frac{2}{r}\frac{dR}{dr} - \frac{n(n+1)}{r^2}R = 0 \tag{13-29a}$$

$$\frac{1}{\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{d\Theta}{d\theta}\right) + \left[n(n+1) - \frac{m^2}{\sin^2\theta}\right]\Theta = 0 \tag{13-29b}$$

$$\frac{d^2\phi}{d\varphi^2} + m^2\phi = 0 \tag{13-29c}$$

where n and m are constants. The solutions of equation (13-29c) are trigonometric functions, and in order that they be single-valued it is necessary that m be an integer. When n and m are both integers, equation (13-29b) has a series of solutions designated by $P_n{}^m(\cos\theta)$.

Problem 38. Show that

$$\Phi = r^n P_n{}^m(\cos\theta)(A\sin m\varphi + B\cos m\varphi) \tag{13-30}$$

is a solution of equation (13-28).

Problem 39. Show that $P_0{}^0 = \text{const}$, $P_1{}^0 = \cos\theta$, and $P_1{}^1 = \sin\theta$ are solutions of equation (13-29b).

Problem 40. Show that

$$\Phi = r^{-(n+1)}P_n{}^m(\cos\theta)(A\sin m\varphi + B\cos m\varphi) \tag{13-30a}$$

is also a solution of equation (13-28).

Problem 41. Show that the functions of the form (13-29) that satisfy equations (13-29a), (13-29b), and (13-29c) form an orthogonal system.

Solutions of Laplace's equation can be obtained by substituting all integral values of m and n such that $m \leq n$. The most general solution is then a linear combination of these particular solutions. In a particular problem it is necessary to evaluate the constants in this linear combination so that the solution will satisfy the prescribed boundary conditions.

Some of the solutions for small values of n and m have simple interpretations.

1. $n = 0$, $m = 0$. The solution corresponding to equation (13-30) is a constant, and that corresponding to equation (13-30a) is $1/r$. This latter is the potential about a point charge at the origin.

2. $n = 1$, $m = 1$. The solution corresponding to equation (13-30) is

$$\Phi = r \sin \theta (A \sin \varphi + B \cos \varphi) \tag{13-31}$$

This represents a uniform field perpendicular to the polar axis and in a direction determined by the relative magnitudes of A and B.

The solution corresponding to equation (13-30a) is

$$\Phi = \frac{\sin \theta}{r^2} (A \sin \varphi + B \cos \varphi) \tag{13-31a}$$

This represents the potential of a dipole perpendicular to the polar axis and in a direction determined by the ratio of A and B.

3. $n = 1$, $m = 0$. Here again the two solutions represent a uniform field and the field of a dipole but in both cases directed along the polar axis.

As an illustration of the method of solution of a general electrostatic problem, consider the case of a dielectric sphere placed in a uniform electric field. The charges producing this field are considered as so far away that they are unaffected by the sphere, and the field at infinity is required to be uniform. The origin may be taken at the center of the sphere, and the potential at the origin may be chosen to be zero. This latter choice is possible because the field at infinity is specified rather than the potential. The potential of the uniform field will then be $\Phi = -E_0 r \cos \theta$. This is one of the spherical harmonic

solutions of Laplace's equation, and it is the only one with a positive power of r that can be used, since it must represent the field at infinity. The effect of the dielectric sphere must then be represented by only those solutions of Laplace's equation that vanish at infinity.

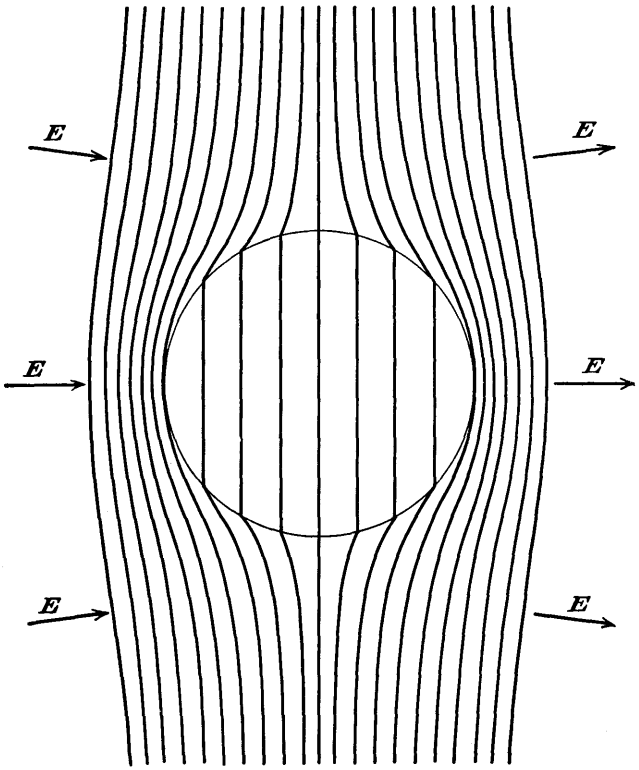


FIG. 13-8.—Some of the equipotential surfaces near a dielectric sphere in a uniform field.

The general procedure would be to assume an infinite series of solutions, with arbitrary constants, and then determine the constants from the boundary conditions. In simple cases, such as the present one, however, it is possible to make a plausible assumption as to the solution and then to see whether or not it is possible to satisfy the boundary conditions. If this can be done, the theorem of uniqueness shows that the assumed solution is correct.

Let us try to represent the potential outside the sphere in the form

$$\Phi_o = -E_0 r \cos\theta + \frac{A \cos\theta}{r^2} \tag{13-32}$$

and inside the sphere in the form

$$\Phi_i = Br \cos\theta \tag{13-32a}$$

The expression for Φ_o contains the term necessary to represent the field at infinity and no other term with a positive power of r. It contains the one term with a negative power of r, and it remains to be shown that the boundary conditions can be satisfied with so simple a solution. The expression for Φ_i contains no negative powers of r since these would lead to an infinite potential of the center of the sphere. Furthermore, since the uniform field has been taken parallel to the polar axis, the potential must be independent of the azimuth angle φ, and only spherical harmonics with $m = 0$ need be used. Considerations of this kind can be used as a guide in selecting the type of function to be used in a trial solution, but the final criterion must be the possibility of satisfying all the boundary conditions.

To evaluate the undetermined constants in the problem at hand we may first apply the condition that the potential is continuous. If the dielectric sphere has the radius R,

$$\Phi_o(R) = \Phi_i(R)$$

and

$$BR \cos\theta = \frac{A \cos\theta}{R^2} - E_0 R \cos\theta$$

or

$$B = \frac{A}{R^3} - E_0 \tag{13-32b}$$

The normal component of the electric displacement must be continuous across the boundary so that

$$K \frac{\partial \Phi_i}{\partial r} = \frac{\partial \Phi_o}{\partial r} \qquad \text{for } r = R$$

or

$$KB \cos\theta = -E_0 \cos\theta - \frac{2A}{R^3} \cos\theta$$

This leads to

$$A = \frac{K-1}{K+2} R^3 E_0 \qquad \text{and} \qquad B = \frac{-3}{K+2} E_0 \quad (13\text{-}32c)$$

and the potentials are therefore

$$\Phi_o = \left(\frac{K-1}{K+2}\frac{R^3}{r^2} - r\right) E_0 \cos\theta \quad (13\text{-}32d)$$

$$\Phi_i = -\frac{3r}{K+2} E_0 \cos\theta \quad (13\text{-}32e)$$

Problem 42. Determine the polarization of the dielectric sphere in the above problem, and use equation (13-14) to find the potential due to it, on the axis and in the equatorial plane.

Problem 43. Find the potential about a conducting sphere in a uniform field.

Problem 44. A conducting sphere is given a charge Q and is placed in a uniform field. Find the potential around it, the distribution of charge on its surface, and the total force on it. The force on each element of area of surface is directed normally outward and is equal to $(\sigma^2/2\kappa_0)ds$.

4. Energy of an Electrostatic System.—For purposes of computing the forces on a few point charges, it is sufficient to use Coulomb's law. For more complicated cases, however, it is more satisfactory to use a potential energy of the system. Of course, this must be clearly distinguished from the potential of the various points.

A single isolated charge experiences no force and so has an energy independent of its position. Let this be set equal to zero here. To bring up another charge while the first is held fixed requires an amount of work equal to $qq'/4\pi\kappa_0 R$, where R is the distance between the two charges. This may be called the energy of the electrostatic system. There is no reason for assigning it to one charge rather than the other, but its negative gradient with respect to the position of either charge gives the force on that charge.

Because the electric field can be derived as the gradient of a scalar potential, it can be shown that the energy of an electrostatic system is a single-valued function of its configuration

and therefore is suitable for use in finding forces. If a series of point charges is brought up from infinity, one at a time, the total amount of work done will be

$$W = \frac{1}{8\pi\kappa_0} \sum_{ij} \frac{q_i q_j}{R_{ij}} \tag{13-33}$$

where both i and i are summed over all the charges. It is clear from the derivation of this expression that the terms for which $i = j$ must be omitted as they have no significance.

Problem 45. Show that the energy of two charges is $qq'/4\pi\kappa_0 R$.

Problem 46. Show that the work necessary to bring a point charge into an electrostatic field is equal to the potential of the point to which it is brought, multiplied by the charge. In this process it may be assumed that all the charges but one are held fixed in position and that only these fixed charges contribute to the potential of the point to which the other charge is brought.

Problem 47. Show that the amount of energy necessary to charge a conductor is $q^2/2C$, where q is the charge put on the conductor and C is its capacitance.

Problem 48. Work out the potential energy of a dipole in an electric field.

When the charges are treated as continuously distributed, the sum in equation (13-33) becomes an integral and has the form

$$W = \frac{1}{2} \int \rho\Phi \, dv \tag{13-34}$$

When it is written in this way there is no difficulty about excluding the potential due to the charge at the point in question, for this goes to zero as dv is made smaller. It does not lead to a singularity as in the case of point charges.

Problem 49. Use equation (13-34) and the solution of Prob. 44 to compute the energy of a charged sphere in a uniform field. From this energy compute the force on the sphere.

To treat the general case it is necessary to compute also the energy associated with the polarization that is present. In this case one must consider the question as to where the zero of

energy is to be put. If the polarization is fixed and its energy of formation is not taken into account, a term

$$-\int \mathbf{P}\cdot\mathbf{E}\,dv = \int \mathbf{P}\cdot\mathbf{grad}\,\Phi\,dv$$

must be included. The more usual case, however, is that in which the polarization is induced and is proportional to the field $\mathbf{E}$ which produces it. In this case the energy stored up in forming the polarization just compensates the negative energy of the dipole in the field that produces it, and the total energy is zero. Hence induced polarization contributes to the energy only in so far as it gives rise to a field at points where charges are located. Under these conditions, when only induced polarization is present, equation (13-34) is clearly the general form for the energy of an electrostatic system.

As an illustration of this situation consider a dipole made up of a positive charge q and a negative charge $-q$, held together by a force proportional to the distance d between them. This force is intended to represent the electrostatic attraction and whatever other forces exist, and may be a good approximation when the distance d is of atomic dimensions. If then a positive charge Q is brought to a distance R from the dipole, the two charges composing it will separate until the force holding them together just balances the force from the charge Q. Under these circumstances d is given by

$$\alpha d = \frac{qQ}{4\pi\kappa_0 R^2} \tag{13-35}$$

where α is the force constant. The work done against this force is $\alpha d^2/2$ so that the energy of formation of this dipole, W_d, is

$$W_d = \frac{1}{2\alpha}\left(\frac{qQ}{4\pi\kappa_0 R^2}\right)^2 \tag{13-35a}$$

The dipole as a whole is attracted by the charge Q, and thus the energy in the field of Q is negative and of amount W_e where

$$W_e = -\frac{qQd}{4\pi\kappa_0 R^2} = -\frac{1}{\alpha}\left(\frac{qQ}{4\pi\kappa_0 R^2}\right)^2 \tag{13-35b}$$

Similarly the energy of the charge Q in the field of the dipole is

$$W_Q = -\frac{Qqd}{4\pi\kappa_0 R^2} = -\frac{1}{\alpha}\left(\frac{qQ}{4\pi\kappa_0 R^2}\right)^2 \qquad (13\text{-}35c)$$

The total energy may be considered as W_d and either W_e or W_Q. In either case

$$W = -\frac{1}{2\alpha}\left(\frac{qQ}{4\pi\kappa_0 R^2}\right)^2 = \frac{1}{2}W_Q \qquad (13\text{-}35d)$$

This is just one-half the energy of Q in the field due to the dipole and is an illustration of the justification for the use of equation (13-34) to express the energy of an electrostatic system containing only induced polarization.

The energy expressed by equation (13-34) is a property of the system as a whole and cannot properly be divided among the charges. This fact is emphasized strikingly by the field point of view when it is applied to this electrostatic energy. It is possible to transform the integral in equation (13-34), which is over the charge density, into an integral over the field.

$$\begin{aligned} W &= \tfrac{1}{2}\int \rho\Phi \, dv = \tfrac{1}{2}\int div\, \mathbf{D}\, \Phi \, dv \\ &= \tfrac{1}{2}\int div\, (\Phi\mathbf{D}) dv - \tfrac{1}{2}\int \mathbf{D} \cdot \mathbf{grad}\, \Phi \, dv \\ &= \tfrac{1}{2}\int \mathbf{D} \cdot \mathbf{E} \, dv \end{aligned} \qquad (13\text{-}36)$$

In this expression the integral of $div\ (\Phi\mathbf{D})$ is set equal to zero because it can be transformed into a surface integral over a surface enclosing the whole system. This surface can be taken so far away that all the field quantities, including Φ and $\mathbf{D}$, vanish rapidly enough to make the integral approach zero. The result is that the energy of the system is one-half the volume integral of the scalar product of $\mathbf{D}$ and $\mathbf{E}$. According to the point of view of Faraday and Maxwell, the energy actually resides in the space surrounding the charges, and one-half the scalar product of $\mathbf{D}$ and $\mathbf{E}$ can actually be regarded as the energy density. This point of view is certainly suggested, although it is not required, by equation (13-36).

There is an apparent discrepancy between equations (13-36) and (13-33) as regards the energy of an isolated charge. In equation (13-33) the terms for which $i = j$ are omitted so that a

single charge is assigned a zero energy. Equation (13-36), on the other hand, would assign it an energy given by the integral over the field. The difficulty is associated with the restrictions it is necessary to place on the size of the volume elements in which charge density is defined. Equation (13-36) leads to no difficulty when the charge distributions are such that a charge density can be used. It cannot be used, however, to compute the field of an isolated electron without further restriction.

Problem 50. Find the energy of a charged conducting sphere by integrating over the electric field outside of it. Compare the result with that obtained by considering the capacitance.

Problem 51. Compute the energy of two charged, concentric, spherical shells.

References

ABRAHAM-BECKER: "Classical Electricity and Magnetism," Chaps. 3 to 5, Blackie & Son, Ltd., Glasgow.

PAGE and ADAMS: "Principles of Electricity," Chaps. 1 to 3, D. Van Nostrand Company, Inc., New York.

PAGE, L.: "Introduction to Theoretical Physics," Chap. 10, D. Van Nostrand Company, Inc., New York.

MASON and WEAVER: "The Electromagnetic Field," Chaps. 1 and 2, University of Chicago Press, Chicago.

CHAPTER XIV

MAGNETOSTATICS AND THE INTERACTION OF STEADY CURRENTS

For many years after the discovery of electricity and the discovery of magnetism, these phenomena were considered to be separate. Although they displayed striking similarities and were cast into the same mathematical form, the studies of magnetostatics and electrostatics were essentially independent. After 1820, when Oersted made the fundamental discovery that an electric current is accompanied by a magnetic field, this situation changed. The two phenomena are no longer considered to be separate. The relationship between them, however, is not associated with their similarities. It is an actual physical relationship such that magnetic phenomena can be produced by electric means and electric phenomena by magnetic means. Oersted's discovery, in fact, emphasized the difference, since there is no means for isolating a magnetic charge or producing a magnetic current. The fundamental magnetic quantity is not charge but magnetization. This is a very fundamental difference and justifies abandoning the early identical treatment for the two subjects and the use of different mathematical descriptions for magnetic and for electric phenomena.

A magnetic field is always produced by an electric current or by magnetization. Magnetization is analogous to electric polarization. It can be crudely pictured as due to molecular currents that act as magnetic dipoles. For this reason the fundamental magnetic phenomenon may be considered to be the interaction between electric currents, and the treatment in this chapter will follow such a line of development. This method of treatment ignores to a large extent some of the striking similarities between electric and magnetic phenomena, but

it emphasizes other more fundamental properties and seems the best introduction to the general theory of the electromagnetic field.

1. Ohm's Law and Steady Currents.—In electrostatic problems a conductor is characterized by the property that all parts of it are at the same potential. This state is attained, however, only after there has been a movement of charge from one part of the conductor to another. Such a movement does not take place instantaneously but requires a certain time, not only because of the inertia of the electrons, but more because the conductor presents a frictionlike resistance to the flow of charge. Such a flow of charge is called a *current*, and the amount of charge passing through a fixed surface per unit time is a measure of the current.

If the two plates of a charged capacitor are connected by a wire, the charge will flow from one plate to the other through the wire and while it is flowing there will be a current in the wire. If the rate at which charge passes a given section of the wire is expressed in coulombs per second, it is a measure of the current in amperes. When expressed in electrostatic units (esu) of charge per second, the current is in esu of current. The measurement of a current requires only the measurement of a charge and a time.

It is found experimentally that, when there is a constant current in a homogeneous wire at constant temperature, the current is proportional to the potential difference between the ends of a section of the wire. The constant of proportionality is called the *conductance* of the section of the wire, and its reciprocal is called the *resistance*. It is found experimentally also that the conductance of a section of wire of uniform cross section s is proportional to s, inversely proportional to the length dl, and proportional to a constant of the material called the *conductivity* σ. Hence

$$I = \frac{-\sigma s}{dl} d\Phi \tag{14-1}$$

where the negative sign indicates that the current has the direction of decreasing potential. The current per unit cross section

of the wire is called the *current density* and may be designated by i. From equation (14-1) it follows that

$$i = \frac{I}{s} = -\sigma \frac{d\Phi}{dl} = \sigma E \qquad (14\text{-}1a)$$

so that the current density is proportional to the electric field in the direction of the wire. The current in a thin wire has the direction of the wire and the sense of the electric field. A current in an extended isotropic conductor has the direction of the electric field, and the current density is a vector proportion to this field.

$$\mathbf{i} = \sigma \mathbf{E} \qquad (14\text{-}1b)$$

If the conductor is not isotropic, the current density is a linear vector function of the field and the conductivity is a tensor rather than a scalar. Only the isotropic case will be treated further in this chapter.

Because the current-density vector represents a flow of electricity, its divergence must be zero in a steady state. Otherwise, charge would be accumulating or disappearing. Except at points where charge is fed into the conductor or taken away from it,

$$div\ \mathbf{i} = 0 \qquad (14\text{-}2)$$

This differential equation describes the current-density vector in an extended conductor. If a vector function can be found that satisfies equation (14-2) and also the boundary conditions and other conditions of the problem, it will be a possible solution of the problem but may not be the only solution. To determine the solution uniquely, other factors may have to be taken into account.

As an example, consider an infinite plane sheet of conductor, and identify points on it by means of a Cartesian coordinate system. Let I coulombs per second of electricity be fed into it at the point $x = -a$ and removed at $x = a$. The problem is thus to find a function whose divergence is zero except at $x = -a$ and $x = a$. The integral of the normal component of $\mathbf{i}$ around any curve enclosing $x = -a$ but excluding $x = a$ must be equal to I.

A vector representing an outward flow from $x = -a$ is

$$\mathbf{i}_1 = \frac{I}{2\pi} \frac{(x + a)\mathbf{i} + y\mathbf{j}}{(x + a)^2 + y^2} \tag{14-3a}$$

Similarly a vector

$$\mathbf{i}_2 = \frac{-I}{2\pi} \frac{(x - a)\mathbf{i} + y\mathbf{j}}{(x - a)^2 + y^2} \tag{14-3b}$$

will represent an inward flow to the point $x = a$. The sum of these is a solution of the problem

$$\mathbf{i} = \frac{I}{2\pi} \left\{ \left[\frac{(x + a)}{(x + a)^2 + y^2} - \frac{(x - a)}{(x - a)^2 + y^2} \right] \mathbf{i} + \left[\frac{y}{(x + a)^2 + y^2} - \frac{y}{(x - a)^2 + y^2} \right] \mathbf{j} \right\} \tag{14-3}$$

In these equations the symbol $\mathbf{i}$ represents both current density and the unit vector along the x axis, but the distinction between these two uses should be clear from the context. Equation (14-3) is not the only solution under the conditions stated. To it could be added a third solution of equation (14-2) such as

$$\mathbf{i}_3 = \frac{-y\mathbf{i} + x\mathbf{j}}{f(x^2 + y^2)} \tag{14-3c}$$

where f is any function of its argument $x^2 + y^2$. This vector has a zero divergence, and its inclusion does not affect the way in which the boundary conditions are satisfied. However $\mathbf{i}_3$ represents a current around the origin, and it is quite evident on physical grounds, and also because of the symmetry, that such a current should not exist in the circumstances of this problem.

The current $\mathbf{i}_3$ can also be eliminated as a solution of this problem by another consideration. Since the electric field in an electrostatic problem can be expressed as the gradient of a potential, its curl must be zero. Hence from equation (14-2) it follows that in an isotropic conductor

$$\mathbf{curl}\ \mathbf{i} = 0 \tag{14-4}$$

If both equations (14-2) and (14-4) are valid at all points, the only possible vector current density is a constant. This means that currents cannot exist in a purely electrostatic field and that

one of the two equations must be inapplicable at some points.

In the above problem, equation (14-2) does not apply at the two points where the charge is introduced and removed, and equation (14-4) does not apply throughout all the circuit leading to and from the plate under consideration.

Problem 1. Consider an infinite solid conductor into which charge is introduced at one point and removed at another. Find the current and the electric field.

According to equation (14-4) the line integral of $\mathbf{i}$ about a closed curve should be zero. It is known, however, that currents do exist in closed circuits, so that such currents must be produced by other than electrostatic means. There must exist electric fields other than the kind described in the previous chapter.

Let $\mathbf{E}'$ be an electric field in the sense that it represents a force on an electric charge, but let it not be the gradient of a potential. The **curl** $\mathbf{E}'$ will not be necessarily zero. This electric field may be due to chemical action as in a battery, to thermal action as in a thermocouple, to electromagnetic action as in a generator, or to other causes. The line integral of the total field $\mathbf{E} = \mathbf{E}_s + \mathbf{E}'$ around a closed circuit will be equal to the line integral of $\mathbf{E}'$ around the circuit and is called the *electromotive force* (emf) of the circuit.

The important distinction between electromotive force and potential difference can be illustrated by the schematic diagram of a circuit fed by a battery as shown in Fig. 14-1. The battery is represented as consisting of two plates marked + and −. The circuit is a long wire of uniform material and cross section so that the resistance per unit length is everywhere the same. Let the potential of the negative pole be zero and the potential of the positive pole be V. The potential will then drop uniformly along the wire. Halfway around the circuit it will be $V/2$.

Inside the battery the potential will likewise fall from V to zero. The field associated with this potential will be $E_s = V/l$, where l is the distance between the plates and this field would tend to stop the current. The characteristic feature of a

battery, however, is that the chemical reaction going on in it tends to move the charge from the negative to the positive pole against the electrostatic field $\mathbf{E}_s$. The energy necessary for doing the work involved comes from the chemical energy of the battery. In Fig. 14-1 the effect of the chemical reaction is represented by the hypothetical electric field $\mathbf{E}'$, and it is clear that $E' > E_s$ in order to overcome the resistance inside the battery.

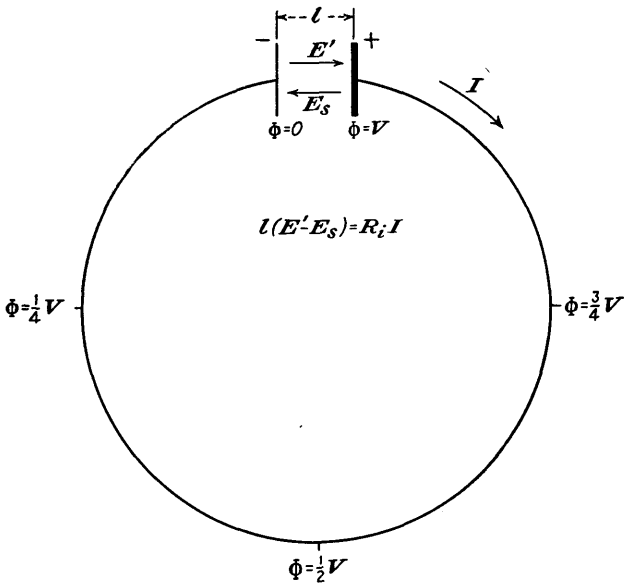


FIG. 14-1.—Potential distribution around a circuit containing a battery.

In this illustration it can be seen that the integral of $\mathbf{E}_s$ around the circuit and through the battery is zero, for the field inside the battery is taken in the opposite direction to that outside. The integral of the whole electric field around the whole circuit is $E'l$, and this is the emf of the battery.

Problem 2. Find the relationship between the emf of a battery and the potential difference between its terminals.

2. Forces between Steady Currents.—As in the case of electrostatics, it is necessary to begin the study of magnetostatics with a fundamental law of force that is the expression of experimental results. This fundamental law seems to have been

first investigated by Ampère about 1825. He studied the forces between closed circuits of various kinds and expressed the results in a series of laws. These laws refer to closed circuits only, since it is not possible to produce a steady current in an open circuit. For the purpose of building a mathematical theory, however, it is desirable to have an expression for the influence of one current element on another. Such an expression can be derived from Ampère's laws, but the result is not unique. There are many laws of force between current elements that will give the observed results for closed circuits. The selection of one of these laws rather than another must be made on the basis of the study of more complex phenomena and by considerations of convenience. The form given below has proved useful in the general theory of the electromagnetic field and is now generally used.

a. The Fundamental Law of Force.—We shall take as the fundamental law of force between elements of a current

$$\mathbf{dF} = \frac{\mu_0}{4\pi} \frac{II'}{R^3} \mathbf{dl} \times (\mathbf{dl}' \times \mathbf{R}) \tag{14-5}$$

This gives the force on the element **dl** due to the presence of **dl′**. **dl** and **dl′** represent elements of thin wires that carry the currents I and I', respectively. The vector **R** is directed from **dl′** to **dl**. As has been indicated, this expression gives the observed forces between closed circuits, but it cannot be directly tested for current elements. Nevertheless, other indications strongly suggest that it is correct. It is also to be especially noted that the expression is not symmetrical in **dl** and **dl′**. This means that the interaction between these two elements does not satisfy Newton's third law of motion and that the momentum of the elements **dl** and **dl′** would not be conserved. The forces between closed circuits, as given by an integral of equation (14-5) as well as observed experimentally, do satisfy this law. Because open circuits must be considered in connection with varying currents, there will be introduced later the idea of momentum and energy existing in the field between the charges and currents. When all the momentum is considered, it will be conserved.

Problem 3. Two elements of current lie in the same plane. Find the force on each of them in terms of the angles they make with the line connecting them.

Problem 4. Find the force per unit length between two long parallel wires of length L that carry currents I and I'.

b. Electromagnetic and Practical Units.—The esu of charge was defined by means of the law of force between two charges. In a similar way, equation (14-5) can be used to define a unit of current. Equation (14-5) gives the force in dynes between two current elements when the current is measured in electromagnetic units (emu) and $\mu_0 = 4\pi$ and is dimensionless. Since current is defined as the amount of charge that passes a given surface per second, the definition of the unit of current gives also the emu of electric charge. The unit of time is the second in all cases. The connection between these emu and the esu of the previous chapter must be determined by experiment, and it will be shown later how this relationship involves the velocity of light.

In using the mksc system of units μ_0 must be given such dimensions and such a value that equation (14-5) holds. Since force is measured in newtons and current in amperes, μ_0 must be measured in newtons per ampere.[2] The value of μ_0 is the result of experiment and is $4\pi \times 10^{-7}$ newtons per ampere[2].

c. The Magnetic Field.—Again as in electrostatics, it is convenient to introduce a vector field to facilitate the calculation of forces. This field is called the *magnetic induction* and is designated by **B**. It is so defined that an element of current put into this field experiences a force given by

$$\mathbf{dF} = I\,\mathbf{dl} \times \mathbf{B} \tag{14-6}$$

The **dF** in this equation differs from the one in equation (14-5) in that it represents the total force on the element **dl** due to the presence of all the other existing currents, while the **dF** in equation (14-5) represents the force due to the current element **dl'** only. From equations (14-5) and (14-6) it follows that

$$\mathbf{B} = \frac{\mu_0}{4\pi} \int \frac{I'}{R^3}\,\mathbf{dl'} \times \mathbf{R} \tag{14-6a}$$

where the integration is carried out over all the elements of current **dl′** which can exert a force on an element placed at the point for which **B** is calculated. The vector **R** points from **dl′** to this point. The vector **B** can be regarded as merely a convenient way of expressing the resultant effects due to all the currents present, and it is often convenient to regard these currents as producing the induction. From equation (14-6) it can be seen that **B** will be measured in newtons per ampere-meter or webers per meter^2.

Problem 5. Show that the magnetic induction around an infinite straight wire is equal in magnitude to $\mu_o I/2\pi\rho$ and is perpendicular to the wire. This can be done by assuming the wire to be finite, computing the field in a plane perpendicular to it at its mid-point, and letting the length increase indefinitely. ρ is the perpendicular distance from the wire to the point at which $\mu_c I/2\pi\rho$ is the induction.

d. The Magnetic Vector Potential.—In electrostatics it is convenient to introduce a potential from which the electrostatic field can be obtained by differentiation, by taking the gradient. With the magnetic induction a similar procedure can be followed, with the difference that the potential is a vector and the induction is obtained by taking the **curl.** In equation (14-6) the vector **R** has its origin at **dl′** and its end at **dl.** The vector $\mathbf{R}/R^3$ can then be expressed as $-$ **grad** $(1/R)$ where the differentiation is with respect to the coordinates of the end of **R,** *i.e.*, with respect to the coordinates of **dl.** This differentiation is independent of the integration over **dl′** and so can be taken out of the integral sign. Then equation (14-6*a*) can be written

$$\begin{aligned}\mathbf{B} &= \frac{\mu_0}{4\pi}\int \frac{I'}{R^3}\,\mathbf{dl'} \times \mathbf{R} \\ &= \frac{\mu_0}{4\pi}\int I'\left[\mathbf{i}\frac{\partial}{\partial x}\left(\frac{1}{R}\right) + \mathbf{j}\frac{\partial}{\partial y}\left(\frac{1}{R}\right) + \mathbf{k}\frac{\partial}{\partial z}\left(\frac{1}{R}\right)\right] \times \mathbf{dl'} \\ &= \frac{\mu_0}{4\pi}\int I'\left[\mathbf{i}\left(\frac{\partial}{\partial y}\frac{dl_z'}{R} - \frac{\partial}{\partial z}\frac{dl_y'}{R}\right) + \mathbf{j}\left(\frac{\partial}{\partial z}\frac{dl_x'}{R} - \frac{\partial}{\partial x}\frac{dl_z'}{R}\right)\right. \\ &\qquad \left. + \mathbf{k}\left(\frac{\partial}{\partial x}\frac{dl_y'}{R} - \frac{\partial}{\partial y}\frac{dl_x'}{R}\right)\right] = \mathbf{curl}\,\frac{\mu_0}{4\pi}\int \frac{I'\,\mathbf{dl'}}{R} \qquad (14\text{-}7)\end{aligned}$$

The **curl** in this equation is with reference to the coordinates of the point at which **B** is desired. This shows that the vector **B** can be derived from a vector potential **A** by taking the curl and that in the case when only steady currents are present

$$\mathbf{B} = \mathbf{curl}\ \mathbf{A} \qquad \text{with } \mathbf{A} = \frac{\mu_0}{4\pi} \int \frac{I\ \mathbf{dl}}{R} \tag{14-8}$$

As an example of the utility of this method, consider the case of a straight wire of length $2L$ carrying a current I. Since all the elements of the wire are in the same direction, the vector potential **A** will have this direction. Take this direction as the z axis of cylindrical coordinates, and compute **A**, in the plane perpendicular to the wire at its center, as a function of ρ and θ.

$$\mathbf{A} = \frac{\mu_0}{4\pi} I\mathbf{k} \int \frac{dz}{(z^2 + \rho^2)^{\frac{1}{2}}} = \left(\frac{2\mu_0}{4\pi}\right) I\mathbf{k} \log \frac{L + (L^2 + \rho^2)^{\frac{1}{2}}}{\rho}$$

As L becomes larger and larger, **A** approaches the value for an infinite wire,

$$\mathbf{A} = \frac{\mu_0}{2\pi} I\mathbf{k} (\log 2L - \log \rho) \tag{14-9}$$

This vector potential has the same dependence on ρ as the electrostatic potential due to a long charged wire worked out in Prob. 3 of the preceding chapter. Although $\log L$ is infinite, it is a constant, independent of ρ, and is immaterial in the determination of **B** from **A**.

The field associated with this vector potential can be found by taking the curl in cylindrical coordinates.

$$\mathbf{B} = \mathbf{curl}\ \mathbf{A} = \frac{\mathbf{e}_\theta(\mu_0/2\pi)I}{\rho} \tag{14-10}$$

This is the same as the result obtained in Prob. 5, with the additional information as to the direction furnished directly by the process of taking the curl.

Problem 6. Find the magnetic vector potential and the magnetic induction about two parallel straight wires of infinite length, both for the case in which the currents are in the same direction and that in which they are in opposite directions.

Problem 7. Show from equation (14-6*a*) that on the axis of a circular loop of wire the magnetic induction is

$$\mathbf{B} = \frac{(\mu_0/2)IA^2}{(A^2 + z^2)^{\frac{3}{2}}}\,\mathbf{k} \tag{14-11}$$

where A is the radius of the loop, z is the distance from the plane of the loop along the axis to the point at which **B** is the induction, and **k** is the unit vector along the axis.

Problem 8. Use the result of the previous problem to show that the induction on the axis of a long, closely wound solenoid is

$$\mathbf{B} = \frac{\mu_0}{2} nI\mathbf{k} \left\{ \frac{L + d}{[(L + d)^2 + A^2]^{\frac{1}{2}}} + \frac{L - d}{[(L - d)^2 + A^2]^{\frac{1}{2}}} \right\} \tag{14-12}$$

where $2L$ is the length of the solenoid, n is the number of turns per unit length, A is the radius, and d is the distance along the axis from the center to the point at which **B** is the induction.

Problem 9. Show that $div\ \mathbf{B} = 0$. (14-13)

Problem 10. Show that the vector potential due to a circular current is

$$\mathbf{A} = \frac{\mu_0}{4\pi} \frac{\mathbf{m} \times \mathbf{R}}{R^3} \tag{14-14}$$

where **R** is the vector from the center of the circle and is much larger than the diameter D of the circle. In this expression **m** has a magnitude given by the product of the current and the area of the circle around which it flows. **m** is perpendicular to the circle, and its direction is related to the direction in which the current is flowing by the right-hand-screw rule. **m** is called the *magnetic moment* of the current.

Problem 11. Find the magnetic induction around the current of Prob. 10, and show that at large distances it has the same form as the electric field around an electric dipole.

c. Current Density.—The fundamental law of force in equation (14-5) was expressed in terms of current elements having a length dl and a cross section so small as to be unimportant. It is sometimes necessary to study the distribution of the magnetic induction in and around large conductors in which the distribution of the current must be taken into account. For this purpose it is necessary to consider the current density.

The connection of the current density with the current is made by means of the equation

$$I\,\mathbf{dl} = \mathbf{i}\,dv \tag{14-15}$$

This equation can be interpreted by taking volume elements in the conductor which are long compared with their transverse dimensions and of which the lengths are parallel to the direction of the current at the point considered. Then each volume element has the properties of a long thin wire to which the left-hand side of the equation can be applied. The right-hand side follows from the definition in equation (14-1*a*). In terms of the current density the vector potential is given by the equation

$$\mathbf{A} = \frac{\mu_0}{4\pi}\int \frac{\mathbf{i}\,dv}{R} \tag{14-16}$$

The integral gives the value of this potential at all points inside and outside of the conductors carrying the current.

3. Properties of the Vector Potential and Magnetic Induction When Due to Steady Currents Only.—From the expression for the vector potential in terms of the current distribution it is possible to determine properties of the field of induction that are sometimes more useful than the defining equations themselves. These properties of the magnetic vectors are analogous to those already described for the electric vectors. It must be remembered that the properties given here refer to the case in which only steady currents are present and do not take into account the presence of material bodies or the existence of variable currents. The magnetic behavior of material bodies will be treated later in this chapter, and the effects of varying currents will be treated in the next chapter.

a. The vector potential is a continuous function of position. This follows from its definition in equation (14-16). Any quantity defined by such an integral will be continuous if the numerator of the integrand is everywhere finite.

b. The vector potential satisfies the partial differential equation

$$\nabla^2\mathbf{A} = -\mu_0\mathbf{i} \tag{14-17}$$

Each component of **A** is given by the defining equation (14-16) in the same form as the scalar electrostatic potential. The Laplacian of this potential was shown to be equal to the negative of the charge density divided by κ_0; it can thus be seen immediately that the Laplacian of each component of **A** is equal to $-\mu_0$ times the corresponding component of **i**. Combining the three component equations gives (14-17).

c. For this special case of steady currents the vector potential satisfies the equation

$$div\ \mathbf{A} = 0 \tag{14-18}$$

This follows from the definition of **A** by taking the divergence

$$div\ \mathbf{A} = \frac{\mu_0}{4\pi}\int\left[i_x\frac{\partial}{\partial x_s}\left(\frac{1}{R}\right)+i_y\frac{\partial}{\partial y_s}\left(\frac{1}{R}\right)+i_z\frac{\partial}{\partial z_s}\left(\frac{1}{R}\right)\right]dv \tag{14-18a}$$

where the subscript s indicates that the (x,y,z) involved are those of the point at which **A** is the potential. The one end of R is at this point, and the other is at the volume element dv. This equation can be written as

$$div\ \mathbf{A} = \frac{\mu_0}{4\pi}\int \mathbf{i}\cdot\mathbf{grad}_s\left(\frac{1}{R}\right)dv = -\frac{\mu_0}{4\pi}\int \mathbf{i}\cdot\mathbf{grad}\left(\frac{1}{R}\right)dv \tag{14-18b}$$

This last equality is based on the fact that

$$\mathbf{grad}_s\left(\frac{1}{R}\right) = -\ \mathbf{grad}\left(\frac{1}{R}\right)$$

as can be seen by writing it out. Because of the vector identity (3) on page 244 it follows that

$$div\ \mathbf{A} = -\frac{\mu_0}{4\pi}\int div\left(\frac{\mathbf{i}}{R}\right)dv + \frac{\mu_0}{4\pi}\int\frac{1}{R}\,div\,\mathbf{i}\,dv \tag{14-18c}$$

Since $div\ \mathbf{i} = 0$ for steady currents, the last integral vanishes. The first can be transformed into a surface integral over any surface bounding the volume containing the currents; and if this surface is taken large enough, the current density on it vanishes. Hence equation (14-18) follows for the case of steady currents with $div\ \mathbf{i} = 0$.

d. From (*b*) and (*c*) it follows that

$$\mathbf{curl\ B} = \mu_0\mathbf{i} \tag{14-19}$$

and from Stokes's theorem that

$$\int_L \mathbf{B} \cdot \mathbf{dl} = \mu_0 I \tag{14-20}$$

where I is the total current through the loop around which the line integral is taken. As will be emphasized later, this relation is valid only in the absence of magnetization.

e. Uniqueness.—If the current density is specified at every point in space, the vector potential can be determined uniquely by means of the definition in equation (14-16). It can also be determined as a solution of equation (14-17) subject to the auxiliary condition (14-18). When determined in this latter way there remains undetermined an additive constant, which, however, is as unimportant as the possible additive constant on the scalar potential.

The proof of this uniqueness follows along lines similar to those in the electrostatic case. If two vectors $\mathbf{A}_1$ and $\mathbf{A}_2$ satisfy the differential equations consider the vector $\mathbf{A}' = \mathbf{A}_1 - \mathbf{A}_2$. Then $div\ \mathbf{A}' = 0$, and from equation (14-17) it follows that $\nabla^2\mathbf{A}' = 0$. This vector equation is the equivalent of three scalar equations, one for each component. Consider A_x'. Then as in equation (13-26),

$$\int_S A_x' \ \mathbf{grad}\ A_x' \cdot \mathbf{ds} = \int (\mathbf{grad}\ A_x')^2\, dv \tag{14-21}$$

The left-hand side of this vanishes if it is required that A_x' vanish at infinity as rapidly as $1/R$. Hence $\mathbf{grad}\ A_x' = 0$ and A_x' must be zero or a constant.

Problem 12. Show that, if equation (14-18) is not taken as the auxiliary condition, there are many different values of the vector potential that will give the correct field of induction, **B**.

Problem 13. Derive equation (14-19).

Problem 14. Use equation (14-20) to find the value of **B** around an infinite straight wire carrying a current.

4. Magnetic Fields Due to Magnetization.—In the previous sections the production of magnetic fields due to steady currents has been treated, but it is also true that there exist magnetic fields around magnetized bodies when there are no currents in

the neighborhood. A magnetized body is not analogous to an electrically charged body, which is accompanied by its electric field; instead, it is analogous to an electrically polarized body surrounded by an electric field characteristic of its state of polarization. A magnetized body is characterized by a magnetic moment rather than by anything like a magnetic charge, and the magnetic moment per unit volume is a vector function of position called the *magnetization* **M**. Magnetic moment has the dimensions of current times area, so magnetization is measured in amperes per meter.

On the molecular picture first developed by Ampère, the magnetic moment of an element of volume is due to closed currents within the molecules. It has been shown above that a small circular current produces a magnetic field which is, at great distances, identical with the field of a dipole. Hence the theory of the field due to magnetization can be taken over almost bodily from electrostatics. For the purposes of a macroscopic theory the detailed nature of the magnetization is unimportant.

In a field due to electric polarization alone, one has from Chap. XIII that, since $div\ \mathbf{D} = 0$, $div\ \mathbf{E} = -(1/\kappa_0)\ div\ \mathbf{P}$, and $\mathbf{curl}\ \mathbf{E} = 0$. Similarly in a field due to magnetization alone

$$div\ \mathbf{H}' = -\ div\ \mathbf{M} \tag{14-22}$$

$$\mathbf{curl}\ \mathbf{H}' = 0 \tag{14-23}$$

Here the field vector is designated by $\mathbf{H}'$. The letter **H** is used instead of **B** because this vector clearly has properties different from those of **B** [equations (14-19) and (14-13)], and the [′] is used to indicate that only magnetization is present. If now we define a new vector $\mathbf{B}' = \mu_0\mathbf{H}' + \mu_0\mathbf{M}$, then

$$div\ \mathbf{B}' = 0 \tag{14-23a}$$

This vector is similar to the previously defined **B** in that it has a zero divergence and can therefore be derived as the curl of a vector potential.

B is the fundamental vector of a magnetic field. When the field is produced by currents only, **B** is given by equations

(14-6) and (14-8). When the field is due to magnetization only, **B** has the properties of the **B′** described above.

In general, both currents and magnetization are present and contribute to the field. In this case the magnetic field vector **B** satisfies the following equations:

$$div\ \mathbf{B} = 0 \tag{14-24}$$

$$\mathbf{curl}\ \mathbf{B} = \mu_0 \mathbf{i} + \mu_0\ \mathbf{curl}\ \mathbf{M} \tag{14-25}$$

$$\mathbf{B} = \mathbf{curl}\ \mathbf{A} \qquad \mathbf{A} = \frac{\mu_0}{4\pi} \int \frac{\mathbf{i} + \mathbf{curl}\ \mathbf{M}}{R}\, dv \tag{14-26}$$

If the distributions of current and magnetization are known, the vector potential can be computed from equation (14-26) and **B** can be found by taking the curl. Equation (14-24) is true for the part of the field due to currents by equation (14-13) and for the part of the field due to magnetization by equation (14-23*a*). The first term on the right side of equation (14-25) comes from equation (14-19) and describes the part of **B** due to currents. The second term applies to the part of **B** due to magnetization only. Similarly the integral for the vector potential contains a term representing the effect of the current and another the effect of the magnetization.

In the general case a vector **H** can be defined by the equation

$$\mathbf{H} = \frac{\mathbf{B}}{\mu_0} - \mathbf{M} \tag{14-27}$$

H is called the magnetic field strength and is often a convenient vector to use. Its curl is equal to the current density and its divergence is equal to $-\ div\ \mathbf{M}$. It may be considered as made up of two independent parts. One part has a zero divergence and a **curl** equal to **i**, and may be computed by means of equation (14-8) if the μ_0 is omitted. The other part has a zero **curl** and a divergence equal to $-\ div\ \mathbf{M}$. It can be computed by regarding $-\ div\ \mathbf{M}$ as an effective magnetic charge and using equations similar to those by which **E** is computed from (ρ/κ_0). The sum of these two parts gives the whole vector **H**.

The difference between **B** and **H** can be illustrated in the case of a long, uniformly magnetized, cylindrical rod in which the magnetization is parallel to the length. Let l be the length

and S the area of the cross section. Under these circumstances the divergence of $\mathbf{M}$ is everywhere zero except at the ends, and the curl of $\mathbf{M}$ is everywhere zero except on the cylindrical surface. On each end of the rod the *div* $\mathbf{M}$ has a large value where the vector rapidly changes from its value of zero outside the rod

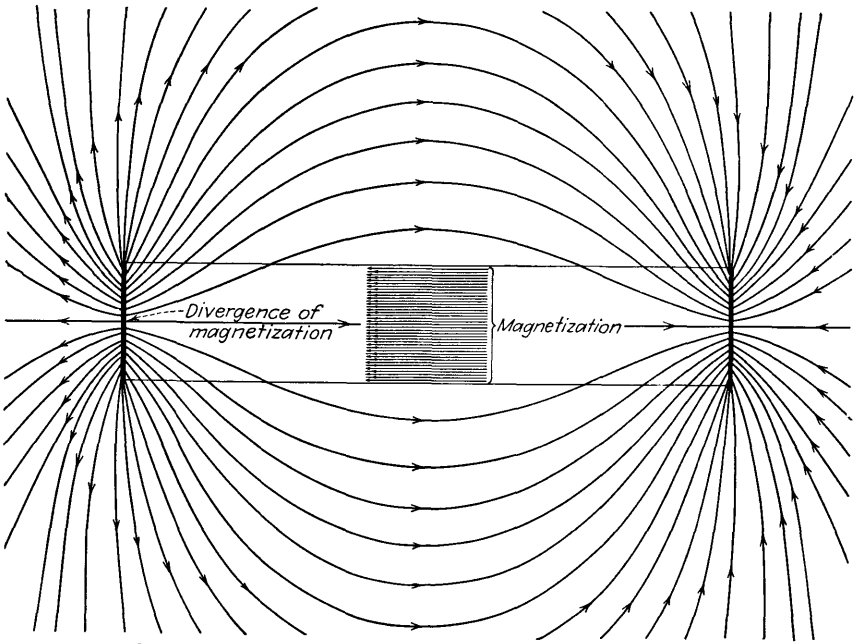


FIG. 14-2.—The field $\mathbf{H}$ around a uniformly magnetized rod.

to the constant value inside. This value of the divergence is such that, when it is integrated over a thin flat volume including the end of the rod, the result is that of a surface integral of the magnitude of $\mathbf{M}$ over the end of the rod. The divergence has values of opposite sign on the two ends. Thus the distribution of $\mathbf{H}$ as shown in Fig. 14-2 is the same as the distribution of the electric field due to two flat surface distributions of charge of surface density M. At the center of the rod $\mathbf{H} = -\frac{8MS}{4\pi l^2}\mathbf{k}$ and is in the direction opposite to the magnetization. Just inside the end at the left of the figure, $\mathbf{H} = -\mathbf{M}/2$ so that at this point

$$\mathbf{B} = \mu_0\mathbf{H} + \mu_0\mathbf{M} = +\mu_0\frac{\mathbf{M}}{2};$$

just outside this end, $\mathbf{B} = \mu_0\mathbf{H} = +\mu_0\mathbf{M}/2$. As a consequence of equation (14-24) the normal component of **B** is continuous across any boundary.

On the other hand, the vectors could have been computed from equation (14-26). The current density is everywhere zero,

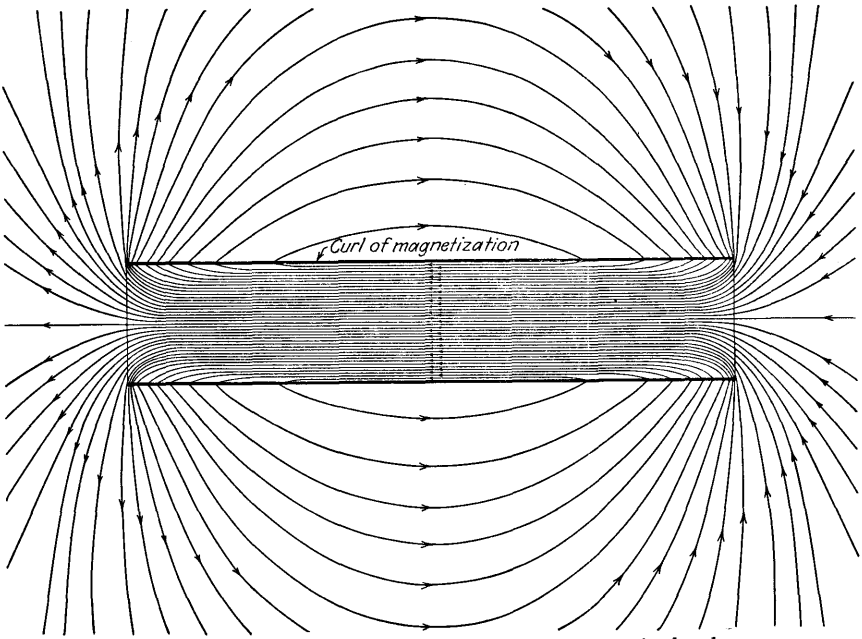


FIG. 14-3.—The field **B** around a uniformly magnetized rod.

but the **curl M** is different from zero in a cylindrical sheet coinciding with the surface of the rod. This curl is parallel to the surface and perpendicular to the axis of the rod, which makes the vector **B** the same as in the case of a long solenoid with a current M per unit length. At the center $\mathbf{B} = \mu_0\mathbf{M}$ if the rod is long enough to neglect the end effects entirely, *i.e.*, long enough to neglect **H** at the center.

Problem 15. Assume that the magnetization at the surface of the rod treated above falls uniformly to zero in a surface layer of thickness t. Compute the value of *div* **M** and **curl M** on the surface.

Problem 16. Determine approximately the values of **B** and **H** inside and around a right circular cylinder of which the radius is much

greater than the length and which is uniformly magnetized parallel to its axis.

As another illustration of the utility of the vectors **B** and **H** and of the application of equations (14-22) to (14-27), consider a large piece of magnetized material in which a cylindrical cavity has been made. Let the axis of the cavity be parallel to the direction of magnetization, and let the radius of the

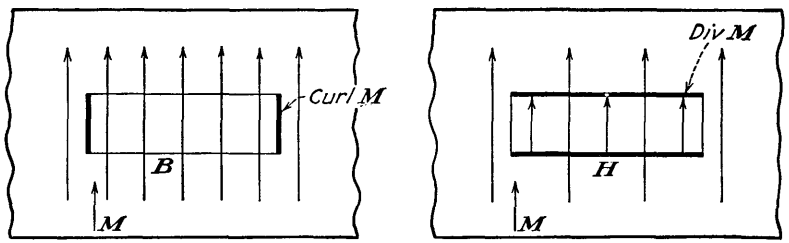


FIG. 14-4.—Values of **B** and **H** inside and around a cavity in magnetic material.

cylinder be much greater than its length. Inside such a cavity it might be possible to measure the magnetic induction **B,** and the question arises as to the connection between this measured **B** and the vectors **B, H,** and **M** in the substance just outside the cavity. It will be assumed that the cavity is so small that its presence will not appreciably affect the magnetic vectors in the material around it and that the field quantities will not change appreciably from one side of it to the other.

From equation (14-24) it follows that the vector **B** will have the same value just inside a circular end of the cylinder as just outside; and since these circular ends are very large compared with the length, one can immediately conclude that the field in the cavity is equal to the **B** in the material. However, it is instructive to consider this on the basis of equation (14-26) as well. The value of **B** inside the material before making the cavity could be computed from the vector potential as determined from equation (14-26). Making the cavity leaves unchanged all the quantities in this expression such as the current density and the **curl M,** with the exception that on the cylindrical wall an additional **curl M** is introduced. The **curl M** introduced along this surface has the same effect on the vector potential as a circular current. The field due to such a current,

at its center, is inversely proportional to the radius of the circle, and therefore in the limit of the case considered here it will be negligible. Thus one concludes again that the insertion of the cavity has no effect on the vector **B**.

On the other hand, if one wishes to work with the vector **H**, the analysis is different but of course leads to the same result. The divergence of **H** is not zero but is equal to $-div\ \mathbf{M}$, and the *div* **M** has a large value on the circular ends of the cavity. On these surfaces the divergence has such a value that in producing a field it is similar to a charge distribution of surface density M. This, as in the electrical case of a flat-plate condenser, produces at the center a field equal to $\mu_0\mathbf{M}$ in the direction of **H** so that the total is $\mu_0\mathbf{H} + \mu_0\mathbf{M} = \mathbf{B}$. The field **H** in this expression refers to the **H** present before making the cavity, since of course in the cavity itself the vectors $\mu_0\mathbf{H}$ and **B** are identical.

Problem 17. Carry through an analysis similar to that above for the case of a spherical cavity, and also for the case of a cylindrical cavity in which the radius is very small compared with the length. Find the relationships between **B** in the cavity and **B**, **H**, and **M**, in the material.

Problem 18. Show that in a field due to both currents and magnetization

$$\int \mathbf{H} \cdot d\mathbf{l} = I \tag{14-28}$$

This is in contrast to equation (14-19), which applies only to the case when no magnetization is present. In that special case, equations (14-19) and (14-28) are identical.

5. Effect of a Magnetic Field on Material Bodies.—When a material body is placed in a magnetic field, magnetization appears in it. This is very similar to the appearance of electric polarization when a body is placed in an electric field. The laws governing the behavior of the material in the field are very complicated. In some cases the magnetization is parallel to the field and proportional to it. The substance is then said to be *paramagnetic*. In other cases the magnetization is not proportional to the field and is very much dependent upon the previous history of the substance. The materials that show the most striking behavior of this kind are in the iron group, and

therefore the behavior is called *ferromagnetism.* In some case the magnetization is opposite in direction to the field. The body is then called *diamagnetic.*

The behavior of paramagnetic and diamagnetic materials can be treated much as was the behavior of a dielectric. The magnetization may be regarded as approximately proportional to the magnetic field so that

$$\mathbf{M} = \chi_m \mathbf{H} \tag{14-29}$$

where the assumed constant of proportionality χ_m is dimensionless and independent of the system of units used. χ_m is positive for paramagnetic and negative for diamagnetic substances.

From equation (14-29) it follows that

$$\mathbf{B} = \mu_0 \mathbf{H} + \mu_0 \mathbf{M} = (1 + \chi_m)\mu_0 \mathbf{H} = \mu \mathbf{H} \tag{14-30}$$

where μ is called the *permeability.* For crystalline bodies the permeability will be a tensor so that the induction is a linear vector function of the magnetic field.

It must be remembered that equations (14-29) and (14-30) are really defining equations for χ_m and μ. Only if χ_m and μ are fairly independent of the magnitude of $\mathbf{H}$ are they useful, but only this case will be considered further.

Problem 19. A toroidal ring of paramagnetic material is closely wound with wire carrying a current. Find the distribution of induction and magnetization throughout the material.

6. Energy in a Magnetic Field.—When dealing with steady currents, the energy of the interaction between circuits may be determined from the fundamental law of force and may be used to express the mechanical forces. The energy can be found by computing the work necessary to bring the circuits into the desired configuration when the currents are kept constant. The work necessary to keep the currents constant, as well as that necessary to start them in the first place, is neglected, so that this result represents only one phase of the magnetic energy.

Consider a closed circuit, carrying a current I, and let

the circuit be at first entirely outside of the region in which the induction **B** due to other circuits has an appreciable value. Let the circuit be turned to be parallel to its desired final position, and then let it be moved, without turning, into this position. Let **dl** be an element of the circuit, and let **dr** be an element of the displacement of the whole circuit. Since the circuit is moved without turning, a single element **dr** describes the motion of all parts of the circuit. Then the element of work done during the displacement **dr** will be

$$dW = -\mathbf{dF} \cdot \mathbf{dr} = -I\int \mathbf{dl} \times \mathbf{B} \cdot \mathbf{dr}$$

This integration is with respect to **dl** and is around the circuit. Thus far there is no integration with respect to **dr**. If the dot and the cross are interchanged, and Stokes's theorem is applied, the result is

$$dW = -I\int \mathbf{B} \times \mathbf{dr} \cdot \mathbf{dl} = -I\int \mathbf{curl}\ (\mathbf{B} \times \mathbf{dr}) \cdot \mathbf{dS}$$

This surface integral is over any surface bounded by the circuit. In taking the curl of the vector product it is to be remembered that **dr** is constant and does not change from one part of the circuit to another. If the curl is expanded, the result is

$$dW = -I\int (\mathbf{dr} \cdot \nabla)\mathbf{B} \cdot \mathbf{dS} = -I\ \mathbf{dr} \cdot \mathbf{grad} \int \mathbf{B} \cdot \mathbf{dS}$$

The total work necessary to move the circuit into the desired configuration is obtained by integration with respect to **dr** and is

$$W = -I\int \mathbf{B} \cdot \mathbf{dS} \qquad (14\text{-}31)$$

This equation shows that there can be defined a kind of potential energy of a circuit in a magnetic field equal to the negative of the product of the current by the total flux of induction through the circuit. This potential energy can be used to determine the force on the circuit. The circuit is urged into such a position that it encircles the maximum flux of induction. It must be emphasized again, however, that this is only the mechanical energy of the circuit. From it the mechanical forces can be derived but the total energy, in which is included the work done in keeping the currents constant, will be treated in the next chapter.

Problem 20. Work out in detail the steps leading to equation (14-31).

Problem 21. A circular wire carrying a current is placed at the center and in the plane of a much larger circular wire through which the same current flows. Find the force on the small coil when it is displaced along the axis from its position of equilibrium.

Problem 22. Find from the energy expression, and then from the fundamental law of force, the restoring torque if the small coil in Prob. 21 is turned about a diameter instead of being displaced.

Problem 23. Show that the energy of a circuit in a magnetic field can be expressed in terms of the magnetic vector potential by

$$W = -I\int \mathbf{A} \cdot \mathbf{dl} \qquad (14\text{-}32)$$

where the integration is around the circuit.

Problem 24. Show from the expression for the energy that the force between two closed circuits satisfies Newton's third law.

CHAPTER XV

THE ELECTROMAGNETIC FIELD

The two previous chapters have treated essentially static situations. The case of a steady current involves motion of the electricity, but the fields produced and the forces exerted are essentially static. These static situations are special cases of the more general case to be treated in the present chapter. To treat this general case of variable currents and fields, it is necessary to take up two extensions of the electric and magnetic equations. These connect the electric and magnetic phenomena very intimately with each other. Before doing this, however, it is of interest to summarize the transition, made in each of the two preceding chapters, between the two basic points of view from which electromagnetic phenomena can be treated. Each chapter began with a point of view from which the interactions between charges and current elements were the objects of attention but ended with the point of view from which the properties of the field itself were regarded as more important. Either point of view is adequate in dealing with the static situations, but for the general case it is necessary to adopt the field point of view.

1. The Electrostatic Field.—This is the special case treated in Chap. XIII. It includes only those situations in which all the electricity is at rest. The results of Chap. XIII may be summarized as follows:

1. The assumed basic law was a law of force between point charges,

$$\mathbf{F} = \frac{1}{4\pi\kappa_0}\frac{q_1 q_2}{R^3}\mathbf{R} \tag{15-1}$$

This was assumed as an expression of experimental facts and was taken for the basis of the rest of the development.

2. A vector field was defined by the relation

$$\mathbf{F} = q\mathbf{E} \tag{15-2}$$

where **F** is the total force experienced by the point charge q at the place where **E** is the electric field. From this definition it follows that the electric field **E** at a given point will be equal to the vector sum of a number of terms of the form (15-1). Through such expressions the field **E** is associated with the charges, which may be said to "produce the field." At this stage **E** is merely a convenient way of expressing the forces between charged particles.

3. An electric potential was defined by the equation

$$\mathbf{E} = -\ \mathbf{grad}\ \Phi \tag{15-3}$$

where Φ is a scalar field quantity. That such a definition is possible is due to the form of the basic law of force, (15-1). The scalar Φ, as well as the vector **E,** is a function of the coordinates, but in the static situations under consideration it is not a function of the time. By means of equations (15-1) to (15-3) the potential Φ is associated with the point charges for which the assumed law of force holds. The scalar Φ is usually more convenient to handle than the vector **E,** but both of them are at first regarded merely as conveniences, defined to help in the application of the basic law of force.

4. It was stated that there exist distributions of positive and negative charge of such a nature as to be best represented as distributions of electric dipoles. The polarization **P** was then defined in terms of the distribution of dipoles throughout a material body, and the vector field **D** was defined by the relationship

$$\mathbf{D} = \kappa_0\mathbf{E} + \mathbf{P} \tag{15-4}$$

In connection with this definition was mentioned the fact that in many practical cases $\mathbf{D} = K\kappa_0\mathbf{E}$ where K is a constant, characteristic of the material, and called the *dielectric constant.*

5. On the basis of the three definitions, (15-2), (15-3), and (15-4), and the assumed law of force (15-1), it was shown that

$$div\ \mathbf{D} = \rho \tag{Is}$$

subject to certain restrictions concerning the possibility of fusing together a large number of point charges into an apparently continuous distribution of charge density ρ. Equation (Is) is a partial differential equation that describes the nature of the field **D**. It no longer emphasizes the "sources" of the field, or the charges that could be said to "produce" the field. The equation is assumed to be true at all points and connects the value of **D** at a given point with its values at neighboring points. This connection is in terms of the charge density ρ at the point in question, but it is only by integrating the equation and evaluating the constants of integration that the field can be expressed in terms of all the charges which produce it.

6. Another equation that follows from the definitions and the assumed law of force is

$$\mathbf{curl}\ \mathbf{E} = 0 \qquad \text{(IIIs)}$$

This is of little use in considering the forces between charges but of great importance in considering the properties of the field itself. Equations (Is) and (IIIs), together with a knowledge of **D** as a function of **E** at every point, serve to describe completely an electrostatic field.

7. Equation (15-3) is the general solution of equation (IIIs). A vector obtained by taking the gradient of any scalar function will satisfy (IIIs), but to satisfy (Is) at the same time it is necessary to find a suitable value for the potential Φ. It was shown in Chap. XIII that this is

$$\Phi = \frac{1}{4\pi\kappa_0} \int \frac{\rho - div\ \mathbf{P}}{R}\, dv \qquad (15\text{-}5)$$

The argument in Chap. XIII was such as to derive this expression from the law of force. It can, however, be derived differently. It can be shown that (15-3) and (15-5) give that particular solution of equations (Is) and (IIIs) associated with a given distribution of charge density ρ and polarization **P**. The only other solutions of the equations are those representing additional fields that are everywhere constant. If the boundary condition is imposed that the fields must not extend to infinity, this trivial generalization is excluded. It is thus possible to

take equations (Is) and (IIIs) as the assumed basic electrostatic laws, instead of assuming the law of force.

2. The Magnetostatic Field.—In the special case of the magnetic fields produced by steady currents, one point of view is characterized by the consideration of the law of force between current elements. In Chap. XIV this point of view was developed to give the properties of the magnetic field vectors. The second point of view, that of the field theory, treats these properties of the magnetic field vectors as the fundamental assumptions. The conclusions of Chap. XIV can be summarized as follows:

1. The law of force between current elements was assumed to be

$$\mathbf{dF} = \frac{\mu_0}{4\pi}\frac{II'}{R^3}\,\mathbf{dl} \times (\mathbf{dl}' \times \mathbf{R}) \tag{15-6}$$

where **dF** is the force on the element **dl** due to **dl'**.

2. A field vector **B**, called the *magnetic induction*, was then defined by the equation

$$\mathbf{dF} = I\,\mathbf{dl} \times \mathbf{B} = \mathbf{i} \times \mathbf{B}\,dv \tag{15-7}$$

where **dF** is the force on the current element I **dl**, or **i** dv, at the point where **B** is the magnetic induction.

3. Because of the properties of this vector field **B**, as expressed in equations (15-6) and (15-7), it was possible to define another field vector **A**, the magnetic vector potential, by the relationship

$$\mathbf{B} = \text{curl}\,\mathbf{A} \tag{15-8}$$

In magnetostatic situations **A** is a function of position, but not of the time.

4. On the basis of the law of force and equation (15-7), or as a consequence of equation (15-8), it follows that

$$div\,\mathbf{B} = 0 \tag{IIs}$$

This property of **B** is the justification for the definition of **A** in equation (15-8), since equation (15-8) is the general solution of (IIs) in terms of the arbitrary function **A**.

5. For use in material bodies another vector, **H**, was defined

by the relationship

$$\mathbf{H} = \frac{\mathbf{B}}{\mu_0} - \mathbf{M} \tag{15-9}$$

where the magnetization **M** was defined in terms of the effective magnetic-moment distribution in the material body. In this connection was mentioned the experimental fact that in many cases $\mathbf{B} = \mu\mathbf{H}$, where μ is characteristic of the material and is called the *magnetic permeability*.

6. It also follows from the law of force that the equation connecting the field quantities with the currents "producing" them is

$$\text{curl } \mathbf{H} = \mathbf{i} \tag{IV}s)$$

Equations (IIs) and (IVs) together with the definitions (15-7) to (15-9) describe the magnetic situation as completely as the law of force (15-6), although they describe it in a different way. They give differential properties of the fields. From them a specific field corresponding to a definite problem can be obtained by integration and insertion of the appropriate boundary conditions. For steady currents a suitable vector potential was shown to be

$$\mathbf{A} = \frac{\mu_0}{4\pi} \int \frac{\mathbf{i} + \text{curl } \mathbf{M}}{R}\, dv \tag{15-10}$$

This leads to fields **B** and **H**, which satisfy (IIs) and (IVs). The procedure of concentrating attention on the field itself is the essence of the treatment of the general electromagnetic field.

3. Electromagnetic Induction.—The electric and magnetic fields described by equations (Is) to (IVs) are quite independent of each other. It is true that the magnetic field is associated with a movement of charge in the form of a current; but with a steady current there is no variation in the electric field, and there may be no significant electric field whatever. Nevertheless, after the discovery of the production of a magnetic field by a current, it seemed natural to look for the production of an electric field by some magnetic means. Because of the lack of magnetic charges and the corresponding magnetic current, it was impossible to produce the exactly analogous situation; and not until the years after 1830 was the phenomenon of electro-

magnetic induction discovered. The discovery was made independently by Joseph Henry in Albany, N.Y., and by Michael Faraday in London. They showed that the results of a variety of experiments could be included in the single statement that a transient current flows in a loop of wire when the flux of magnetic induction through the loop changes. This can be formulated in the equation

$$I\mathfrak{R} = -\frac{d}{dt}\int \mathbf{B}\cdot\mathbf{dS} \qquad (15\text{-}11)$$

In this equation I is the current in the wire, measured in amperes. It is not necessarily a steady current but flows just during the time in which the flux of magnetic induction is changing. $\mathfrak{R}$ is the resistance of the wire, and $\mathbf{B}$ is the magnetic induction. The integral is taken over any surface bounded by the circuit, and the positive direction of I is related to the direction of the surface element $\mathbf{dS}$ by the right-hand rule used in Stokes's theorem. The equation as given is correct for quantities measured in the practical system. If the quantities on the left side of equation (15-11) are measured in esu and those on the right in emu, the dimensions of the two sides are the same only if the right-hand side is divided by a velocity. Experiment shows that this must have the magnitude of the velocity of light in vacuum, so the equation holds if the right-hand side is multiplied by $1/c$.

In the previous chapter the product of the current flowing in a wire by the resistance between two points of it was shown to be equal to the potential difference between those two points. However, in the case of a closed loop, there can be no potential difference around the whole loop, since the potential is a single-valued function of position. The effect of the change in the flux of magnetic induction must then be described as an emf. This emf is the integral, around the loop, of an electric field. It is an electric field, however, that cannot be described as the gradient of a potential and that does not satisfy equations (15-3) and (IIIs). If E' represents the strength of this induced electric field, equation (15-11) is equivalent to

$$\mathcal{E} = \int_L \mathbf{E}'\cdot\mathbf{dl} = -\frac{d}{dt}\int_S \mathbf{B}\cdot\mathbf{dS} \qquad (15\text{-}12)$$

In so far as equation (15-12) is identical with equation (15-11), the line integral is to be taken around a loop of wire; but since the tangential component of an electrostatic field was found to be continuous across the boundary between two substances, it is natural to regard (15-12) as applicable also to a path just outside the wire. After that it is not difficult to think of the equation as applicable even after the wire is removed and thus as applicable to any closed path whatever.

If equation (15-12) is assumed to apply to any closed path, it is convenient to use Stokes's theorem to write the corresponding differential equation

$$\text{curl } \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \qquad \text{(III)}$$

E is written here instead of **E′**, since the curl of any electric field due to fixed charges is zero. This **E** then refers to the field produced by electric charges plus the field produced by electromagnetic induction. With the restriction to fields of these types, equation (III) will replace equation (IIIs), which was valid in static situations only. In the static cases (III) reduces to (IIIs). Equation (III) can be regarded as the differential form of the law of electromagnetic induction, although it contains in some respects more and in other respects less than the results of the original experiments.

Equation (III) contains more than the original experimental results in that it implies the existence of an electric field without regard to the presence or absence of a conductor. It implies that an electric charge will experience a force in the presence of a changing magnetic field, and in this respect it is a true generalization of the experimental results. This generalization will now be taken as a basic assumption instead of equation (IIIs), and it thus becomes one of the postulates of the general theory. In recent years this assumption has been the basis of construction of a type of electron accelerator known as the *Betatron*.

On the other hand, equation (III) contains less than the results of the original experiments in that it does not describe the emf induced in a circuit when the circuit itself is moved in the magnetic field. The flux of induction through a loop can

be changed either by moving the loop to a place where the magnetic induction has a different value or by changing the value of the induction while keeping the circuit fixed. Both these processes are envisaged by equation (15-11), but only the latter is covered by equation (III). For most problems equation (15-11) can be used directly to include both cases, but as a general basic hypothesis equation (III) must be adopted. The method of treating cases in which the circuits are moved will be taken up in the next section.

Problem 1. A circle of wire is placed inside a long solenoid with its plane perpendicular to the axis of the solenoid. If the current through the solenoid is equal to $I_0 \sin \omega t$, find the emf in the wire. Neglect the magnetic field due to the current flowing in the wire itself. This is a good approximation in cases where the resistance of the wire is high.

Problem 2. Treat the above problem when the current in the loop is not neglected. Show that

$$\frac{d\mathcal{E}}{dt} + \frac{\mathfrak{R}}{L}\mathcal{E} = -\frac{\mu_0 n\omega R}{L} A I_0 \cos \omega t \qquad (15\text{-}13)$$

where $\mathcal{E}$ is the emf around the loop, $\mathfrak{R}$ is its resistance and A its area, n is the number of turns per unit length of the solenoid, and L is a constant such that the total flux of induction through the loop due to a current I flowing in it is LI. L is called the *coefficient of self-induction* of the loop. Also, find the current in the loop.

4. Fields in Moving Coordinate Systems.—The vector **E** in the equations of the electromagnetic field refers to the electric field at a given point and time, (x,y,z,t), specified with reference to a given set of coordinate axes. The person observing and describing the phenomena is regarded as at rest with reference to these axes. Suppose, however, that the same phenomena are to be observed and described by another person, who wishes to use coordinate axes moving with respect to the first. In these axes let the corresponding coordinates and time be (x',y',z',t'). This second observer may be considered to be at rest with reference to this second set of axes, and it is assumed to be just as possible to describe the phenomena from his point of view as from the point of view of the first observer. The

question then arises as to whether the two observers will describe the phenomena in the same way or whether they will differ in essential respects.

It is possible to write immediately the connection between the coordinates and the time (x,y,z,t), of an event as measured with reference to the first set of axes and clocks, and the coordinates of the same event, and the time, (x',y',z',t'), as measured with reference to the second set of axes and clocks. To avoid any relativistic effects it will be assumed that the relative velocity v of the two sets of axes is constant and very small compared with the velocity of light. Under these circumstances the transformation equations for the coordinates and the time are

$$\left.\begin{aligned} x &= x' + v_x t \qquad y = y' + v_y t \\ z &= z' + v_z t \qquad t = t' \end{aligned}\right\} \tag{15-14}$$

The problem is then to find out whether the fields **E** and **B** at a given point are the same for both systems of axes or whether they appear differently to the two observers.

Let S be the set of axes with reference to which (x,y,z,t) are the coordinates and the time of an event, and let S' refer to the other set, which is moving with the constant velocity **v** with reference to S. Assume the observer at rest in the system S finds that at the point P there is no electric field but that there is the magnetic induction **B.** This he would find by placing at the point P a stationary electric charge q and noting that it experiences no acceleration. On the other hand, if he moves the charge with the velocity **v,** he finds it subject to such an acceleration as to indicate the action of a force equal to $q\mathbf{v} \times \mathbf{B}$. Now let the observer associated with the axes S' perform similar experiments. He places at the point P a charge that to him is stationary. This means that it is stationary with respect to the axes S' but that with reference to the axes S it is moving with the velocity **v.** The observer in S would see such a charge accelerated as by a force $q\mathbf{v} \times \mathbf{B}$, and the observer in S' will see it accelerated by the same amount. That the accelerations, and consequently the forces, as seen by both observers are the same follows from the form of the transformation equations

(15-14) for the coordinates and the time. Since the observer in S' is using a charge that to him is stationary, he must conclude that there exists at the point P an electric field of magnitude $\mathbf{v} \times \mathbf{B}$. This is clearly in addition to the force due to any electric field which might have been observed by the observer in S, so that the fields as seen by the two observers are connected by the transformation equation

$$\mathbf{E}' = \mathbf{E} + \mathbf{v} \times \mathbf{B} \qquad (15\text{-}15)$$

$\mathbf{E}'$ is the electric field with reference to the axes S', and $\mathbf{E}$ and $\mathbf{B}$ are the electric field and magnetic induction, respectively,

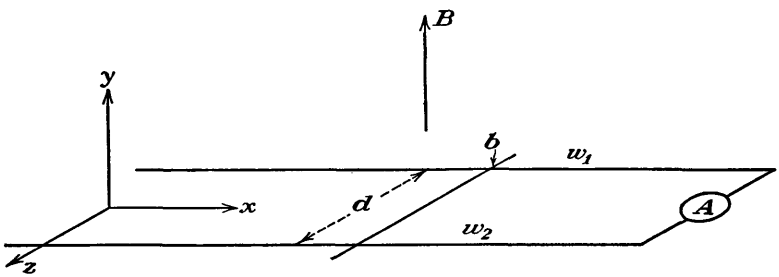


FIG. 15-1.—A sliding bridge across two wires in a magnetic field.

with reference to the axes S. It is to be emphasized that these fields are at the same point and at the same time. The difference is entirely in the system of axes with reference to which they are described.

One important use of the transformation of fields is in the determination of the emf in a moving conductor. In the preceding chapter it was stated that the current density is equal to the electric field multiplied by the conductivity. Since only steady currents were under consideration it was not necessary to state that the field was referred to a system of axes with reference to which the conductors were stationary; but such a specification was implied. It is possible to complete the definition of emf in a conducting circuit by stating it to be the line integral, around the circuit, of the electric field, where, at each point, the field is measured with reference to axes moving with the conductor.

As an illustration of the application of this idea, consider the arrangement shown in Fig. 15-1. Two straight wires w_1

and w_2 are parallel to the x axis, in the x-z plane, and a distance d apart. They are connected together at one end through an ammeter, which makes possible the measurement of the emf when the resistance is known. Another short wire b is parallel to the z axis and makes contact with w_1 and w_2. The whole arrangement is in a uniform magnetic field **B**, parallel to the y axis. Let the connecting wire, b, be moved with constant velocity v in the direction of increasing x, and calculate the emf of the circuit. There are three principal methods by which this is usually done.

1. By means of equation (15-12) the emf can be equated to the rate of change of the area of the circuit multiplied by the normal component of the induction **B**. The rate of change of the area is vd so that this procedure gives

$$\mathcal{E} = vdB$$

This is a convenient method for many cases. It will always work, as will be shown later, when the circuit is permanently connected together without sliding contacts. In the above case it works even in the presence of sliding contacts.

2. The induced emf is often said to be equal to the rate at which lines of flux are cut by the moving parts of the circuit. The number of lines of flux per unit area is defined as the component of **B** normal to the surface. The number of lines of flux cut per unit time in the above arrangement is clearly vdB, so that this method gives the same result as the previous one. It is necessary, of course, to specify some means of determining the sign of the emf but this can be done easily. This method may lead to confusion when the source of the magnetic induction is not stationary but is also moving.

3. The third method makes use of the transformation equation for the fields. Along the wires w_1 and w_2 there is no electric field and no contribution to the emf. Along the wire b, however, the field has a value, which must be obtained from equation (15-15). To an observer moving with the wire there is an electric field of magnitude $E_z' = vB$, and the emf is just this quantity multiplied by the distance d along which it acts. Clearly this gives the same result as the other two methods.

In the above example all three methods give the same result, and one might conclude that they should be regarded as equally satisfactory. The reason for adopting the last method as the most fundamental is that it permits the treatment of problems to which the first two methods cannot be applied unambiguously. The following problems are illustrative of such cases:

Problem 3. A very long conducting strip of thickness t and width d lies in the plane normal to a uniform field of magnetic induction **B.** At two points on opposite edges of the strip are sliding contacts, which connect the edges together through a circuit containing an ammeter.

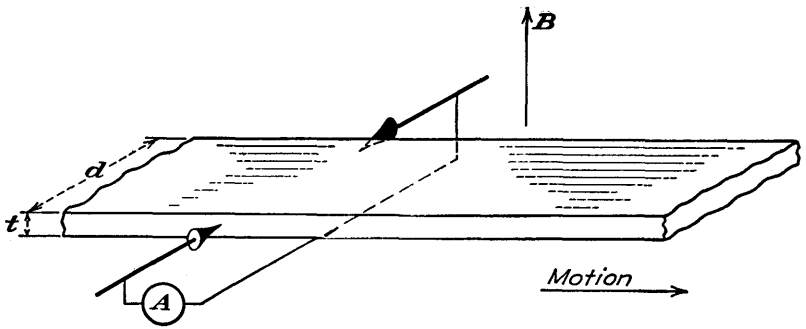


FIG. 15-2.—Arrangement for inducing an electromotive force by moving a conducting strip parallel to its length.

The strip is then moved parallel to its length with the constant velocity v. What current, if any, will flow through the ammeter? The resistance of the complete circuit is $\mathcal{R}$ (see Fig. 15-2).

Problem 4. Compute the emf induced in a solid metal wheel rotating about its axis which is parallel to a uniform magnetic induction. The emf can be measured by the current flowing in a wire that makes sliding contact at the edge and at the axis of the wheel.

It is not difficult to show that this method of transformation of the fields, applied to a moving circuit or to a circuit that is being deformed in an unchanging field, leads to equation (15-12). Let **v** be the vector velocity of the element of the circuit **dl.** Then, by equation (15-15),

$$\begin{aligned} \int \mathbf{E} \cdot \mathbf{dl} &= \int \mathbf{v} \times \mathbf{B} \cdot \mathbf{dl} = -\int \mathbf{B} \times \mathbf{v} \cdot \mathbf{dl} \\ &= -\int \mathbf{B} \cdot \mathbf{v} \times \mathbf{dl} \end{aligned} \tag{15-16}$$

The last integral is just the rate of change of the area of the

circuit multiplied by the magnetic induction, and thus equation (15-16) is equivalent to (15-12) when **B** does not change with the time. When both kinds of changes are taking place, when the induction is changing with the time and the circuit is also being moved or deformed, the emf will be composed of two parts. One of these will be based on equation (15-16) and the other on equation (III). This leads to

$$\int \mathbf{E} \cdot \mathbf{dl} = -\left\{ \int \frac{\partial \mathbf{B}}{\partial t} \cdot \mathbf{dS} + \int \mathbf{B} \cdot \mathbf{v} \times \mathbf{dl} \right\}$$
$$= -\frac{d}{dt} \int \mathbf{B} \cdot \mathbf{dS} \qquad (15\text{-}17)$$

This indicates that in the general case, when a simply connected circuit without sliding contacts is under consideration, equation (15-12) will give the induced emf correctly. The restriction to a particular kind of circuit is due to the equating of $\int \mathbf{B} \cdot \mathbf{v} \times \mathbf{dl}$ to the rate of change of flux through the circuit due to its motion or deformation. In Probs. 3 and 4 such a possibility did not exist.

Problem 5. A circular loop of wire is turned with constant angular velocity about a diameter as an axis. It is in a magnetic field that is uniform but that changes sinusoidally with the time. This change has a frequency different from the rate at which the coil is being turned. Find the emf induced in the coil when the current flowing in it is neglected.

5. The Energy in a Magnetic Field.—In the preceding chapter it was shown that the mechanical work necessary to bring a circuit carrying the current I into a magnetic field is given by

$$W = -I \int_S \mathbf{B} \cdot \mathbf{dS} \qquad (15\text{-}18)$$

where the integral is taken over any surface bounded by the circuits. The vector **B** as used in this expression refers only to that part of the magnetic induction independent of the current I in the circuit itself. The negative sign indicates that the forces are such as to pull the circuit into a position in which the flux through it is a maximum.

Equation (15-18) was derived on the assumption that all currents were kept constant while the relative positions of the circuits were being changed, and no account was taken of the means by which this was to be done. No account was taken of the phenomenon of electromagnetic induction. When account is taken of this, of the fact that the work per unit time necessary to keep a current I flowing against an opposing emf is $-I\mathcal{E}$, it is possible to compute from equation (15-11) the work necessary to keep the currents constant while the circuits are being moved. For simplicity consider only two circuits, 1 and 2 (a larger number could be included without any essential change in the argument). Let W_1 be the work done in keeping up the current I_1 in circuit 1 while it is moved from a position in which the flux of induction is zero up to the desired position. Then

$$\frac{dW_1}{dt} = -I_1\mathcal{E}_1 = I_1 \frac{d}{dt} \int_S \mathbf{B}_2 \cdot \mathbf{dS}_1$$

whence

$$W_1 = I_1 \int_{S_1} \mathbf{B}_2 \cdot \mathbf{dS}_1 \qquad (15\text{-}19)$$

where $\mathbf{B}_2$ is the induction due to the current I_2 in circuit 2. Similarly

$$W_2 = I_2 \int_{S_2} \mathbf{B}_1 \cdot \mathbf{dS}_2 \qquad (15\text{-}19a)$$

Careful attention must be given to the signs in these equations. The statement that the work done per unit time is equal to $-I\mathcal{E}$ is based on the convention by which both I and $\mathcal{E}$ are taken as positive in the same sense around the circuit. This sense is related to the positive direction of the surface elements $\mathbf{dS}$ by the rule used in Stokes's theorem.

The total work done in bringing the circuits together is the sum of W_1, W_2, and the mechanical work of equation (15-18). It can be shown by using the expression for $\mathbf{B}$ in terms of the vector potential and the expression for $\mathbf{A}$ in terms of the currents that

$$I_1 \int_{S_1} \mathbf{B}_2 \cdot \mathbf{dS}_1 = I_2 \int_{S_2} \mathbf{B}_1 \cdot \mathbf{dS}_2 \qquad (15\text{-}20)$$

From this it follows that all the three components of the work

are of the same magnitude but two are positive in sign and one is negative. The result is that the total work done is given by

$$W = I\int \mathbf{B} \cdot \mathbf{dS} \tag{15-21}$$

where the current and the integral refer to either circuit but the induction is that due to the other circuit only. This quantity may be called the *mutual electromagnetic energy* of the pair of circuits. Another way of writing this would be to attribute half the energy to each circuit, but this would have as little significance as any other division of mutual energy among the component parts of a system.

If we consider only cases in which the magnetic permeability is not a function of the magnetic field but may still be a function of position, the magnetic induction at any point may be divided into parts due separately to each of the circuits present. Each of these parts will be proportional to the current flowing in the circuit that produces it, according to equation (14-6). This fact makes it possible to regard the flux of induction through a circuit as a sum of terms proportional to the various currents involved.

$$\int_{S_i} \mathbf{B} \cdot \mathbf{dS}_i = L_i I_i + \sum_{j \neq i} M_{ij} I_j \tag{15-22}$$

L_i is called the *coefficient of self-inductance* of circuit i and M_{ij} is called the *coefficient of mutual inductance* between circuits i and j. L_i is the flux of induction through circuit i when unit current is flowing in it and no other currents are in the neighborhood. Similarly, M_{ij} is the flux of induction through circuit i due to a unit current in circuit j. These coefficients depend upon the shapes and the relative positions of the circuits and upon the distribution of permeability in the surrounding material. Since the flux of magnetic induction is measured in webers, the coefficients of induction are measured in webers per ampere, or henrys.

Problem 6. Show that $M_{ij} = M_{ji}$, and justify equation (15-20).

Problem 7. Calculate the mutual inductance between a long solenoid and a short coil placed inside of it.

Problem 8. Calculate the mutual inductance between a large

circular coil and a small circular coil, when they are coaxial and lie in parallel planes.

A system of circuits can be brought into a desired situation in two characteristic ways, for which the necessary work can be calculated. One way is to start the current in each of the circuits and to bring it up to its desired value while they are all far apart and do not influence each other and then to bring the circuits into their desired positions while the currents are kept constant. The other way is first to put the circuits into the desired positions and then to increase all the currents together from zero.

Consider first the former method. The energy of the final configuration of circuits will be equal to the work necessary to start the currents flowing in each circuit plus the work necessary to bring the circuits together. The work necessary to start a current in one circuit is determined by the differential equation

$$\frac{dW_s}{dt} = -I\mathcal{E} = I\frac{d}{dt}\int_S \mathbf{B}\cdot\mathbf{ds} = LI\frac{dI}{dt}$$

whence

$$W_s = \tfrac{1}{2}LI^2 \tag{15-23}$$

For the work necessary to bring the circuits together a term of the form (15-21) appears for each pair of circuits. The whole energy is then

$$W = \tfrac{1}{2}\sum_i LI_i^2 + \tfrac{1}{2}\sum_{j\neq i} M_{ij}I_jI_i \tag{15-24}$$

This expression is useful when the system is composed of a small number of well-defined circuits. When the distribution of current density in large conductors is involved, other forms are more convenient.

Problem 9. Show that the work done in changing one current while holding the others constant is correctly given by the partial derivative of W with respect to the current in question, multiplied by the change in the current. Part of the work is necessary to hold the currents constant when the induction is changed.

Problem 10. Derive equation (15-24) by the second of the methods mentioned above.

Problem 11. Consider a circuit of resistance $\mathcal{R}$ and self-inductance L. Let there be applied to it an emf given by $\mathcal{E}_0 \sin \omega t$, and compute the current.

Problem 12. Consider two circuits of resistances $\mathcal{R}_1$ and $\mathcal{R}_2$, self-inductances L_1 and L_2, and carrying currents I_1 and I_2. Let the mutual inductance be M_{12}. If an emf $\mathcal{E}_0 \sin \omega t$ is applied to circuit 1, show that

$$L_1 \frac{dI_1}{dt} + M_{12} \frac{dI_2}{dt} + \mathcal{R}_1 I_1 = \mathcal{E}_0 \sin \omega t \tag{15-25a}$$

$$L_2 \frac{dI_2}{dt} + M_{12} \frac{dI_1}{dt} + \mathcal{R}_2 I_2 = 0 \tag{15-25b}$$

The expression (15-24) for the energy of a system of circuits is written from the point of view of the currents and the forces they exert on each other. The energy depends on the magnitudes of the currents and, through the coefficients of inductance, on the position of each element of current. The energy can also be expressed, however, as an integral over the field vectors. Equation (15-24), as can be seen from the method of derivation, is equivalent to one-half the sum, over all the circuits, of the current in each one multiplied by the total flux of induction through it. This includes the induction due to the current in the circuit itself. Thus

$$W = \tfrac{1}{2} \sum_i I_i \int_{S_i} \mathbf{B} \cdot \mathbf{dS}_i = \tfrac{1}{2} \int \mathbf{A} \cdot \mathbf{i}\, dv = \tfrac{1}{2} \int \mathbf{A} \cdot \mathbf{curl}\, \mathbf{H}\, dv$$

$$= \tfrac{1}{2} \int \mathbf{B} \cdot \mathbf{H}\, dv \tag{15-26}$$

The sum in the first expression is over all the circuits. When it is replaced by a volume integral of the current density, this integral must be taken over all space. Since by means of equation (15-26) it is possible to obtain the total energy of the system of circuits by an integration over all the magnetic field, one is tempted to regard the energy as located in the field and to say that $(\mathbf{B} \cdot \mathbf{H})/2$ is the density of energy. This conception of a field energy distributed throughout the space in which there is a magnetic field is indispensable in the treatment of problems of radiation, but it is not required by the derivation of equation (15-26).

Problem 13. Work out in detail the steps in equation (15-26).

Problem 14. Find the electromagnetic energy of a long solenoid carrying a current, on the assumption that the field is uniform inside it and zero outside it.

6. The Displacement Current.—The second extension of the fundamental equations necessary to treat the general case is concerned with curl of the magnetic induction. According to the equation for the static case

$$\mathbf{curl}\ \mathbf{H} = \mathbf{i} \qquad \text{(IV}s\text{)}$$

Because the divergence of a curl is identically zero, this equation implies that the divergence of the current is always zero. While this is true for the case of steady currents, to which equation (IV*s*) is applicable, it is not true in the more general case. If a condenser is being charged through a wire, the current ends at the condenser plate and its divergence is not zero at this point. However, the divergence of the current is necessarily accompanied by a disappearance of charge, and the general principle of the conservation of charge is expressed by the equation

$$div\ \mathbf{i} + \frac{\partial \rho}{\partial t} = 0 \qquad \text{(15-27)}$$

Since $div\ \mathbf{D} = \rho$, the quantity $\mathbf{i} + \partial\mathbf{D}/\partial t$ is a vector whose divergence is zero. If equation (IV*s*) is now generalized to

$$\mathbf{curl}\ \mathbf{H} = \mathbf{i} + \frac{\partial \mathbf{D}}{\partial t} \qquad \text{(IV)}$$

the divergence of the left side is identically zero, and the divergence of the right side is zero because of the conservation of charge. The equation is thus consistent under all circumstances. The quantity $\partial\mathbf{D}/\partial t$ is the *displacement current density.* Its inclusion is often regarded as the outstanding contribution of Maxwell to the electromagnetic theory. This generalization was made, not as the result of experiments, but rather for reasons such as those just outlined. Nevertheless, subsequent experiments have provided full justification for its inclusion as one of the fundamental hypotheses of the electromagnetic theory.

There is a marked similarity between equation (IV) and equation (III), but the absence of a magnetic conduction current in equation (III) constitutes an essential difference. This is paralleled by the corresponding difference between equations (I) and (II) due to the absence of any magnetic charge density.

In many problems the displacement current is negligible when compared with the conduction current. This is true even with rather rapidly alternating currents. However, when the frequency becomes very high, as in problems dealing with radio transmission, the displacement current cannot be neglected. In fact, the displacement current may be regarded as principally responsible for the phenomena of electromagnetic waves.

Problem 15. The two plates of a condenser are connected by a wire. Show that when there is a current in the wire the total displacement current between the condenser plates is equal to the conduction current in the wire.

Problem 16. Find the magnetic field due to a charged particle moving in a straight line with a constant velocity small compared with the velocity of light. This involves consideration of the rate of change of the electric field.

7. Maxwell's Equations.—We may now collect the four equations that are to be regarded as the basic equations of the electromagnetic field. These are usually called *Maxwell's equations* because of Maxwell's work in developing the mathematical formulation of the theory. The four field equations are as follows:

$$\left.\begin{aligned} &\text{(I) } div\,\mathbf{D} = \rho \qquad \text{(II) } div\,\mathbf{B} = 0 \\ &\text{(III) } \mathbf{curl}\,\mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \qquad \text{(IV) } \mathbf{curl}\,\mathbf{H} = \mathbf{i} + \frac{\partial \mathbf{D}}{\partial t} \end{aligned}\right\} \tag{15-28}$$

To the field equations must be added the force equation, which permits a definition of the field quantities in terms of the mechanical concept of force.

$$\mathbf{F} = \rho\mathbf{E} + \mathbf{i} \times \mathbf{B} \tag{15-29}$$

$\mathbf{F}$ is the force per unit volume on an element of volume in which the charge density is ρ and the current density is $\mathbf{i}$.

It is also necessary to include in the list of the fundamental equations the relationships between **E** and **D** and between **B** and **H**. For it is through them that the effects of electric and magnetic dipoles are included in Maxwell's equations. These relationships are

$$\mathbf{D} = \kappa_0\mathbf{E} + \mathbf{P} = \kappa\mathbf{E} \qquad \mathbf{B} = \mu_0\mathbf{H} + \mu_0\mathbf{M} = \mu\mathbf{H} \quad (15\text{-}30)$$

The expressions of **D** in terms of **E** and **P** and of **B** in terms of **H** and **M** are generally true by definition, but the use of the dielectric permittivity κ and the magnetic permeability μ is only generally valid if these quantities are not required to be constants but are made functions of **E** and **H**, respectively. If they are not constants, their usefulness is small. In the present work we shall consider only the simple cases in which μ and κ are constants for a given piece of material. This excludes the cases in which there is permanent magnetization, magnetic or electric hysteresis, or other similar phenomena that are to a greater or less degree characteristic of all real materials.

8. Units.—It is important to note carefully the units used in writing these equations. Here, as in the previous chapters, the practical, or mks, system has been used because of its recent widespread adoption.

In reading earlier papers on electricity and magnetism other systems of units will be encountered. One of the most prevalent of these is the Gaussian system. In this system the electric quantities, electric field, displacement, charge, current, etc., are measured in the esu mentioned in Chap. XIII. The magnetic quantities such as magnetic induction, magnetic field, magnetization, etc., are measured in the electromagnetic system mentioned in Chap. XIV. Since the Maxwell equations contain both electric and magnetic quantities, the equations themselves must define a relationship between these two systems of units. This connection is given by the value of a quantity c, which is the velocity of electromagnetic waves in a vacuum. In the mks units, this velocity is implicit in the quantities κ_0 and μ_0.

The following table gives the relationship between the units of a number of quantities in the mks, or practical, system, the electrostatic system, and the electromagnetic system. The

electrostatic and electromagnetic are designated by the subscripts s and m, respectively. The ratios are expressed in terms of the measures of a fixed quantity in the indicated units. They are not the ratios of the sizes of the units themselves. For example, if a given current is I amperes, the same current will be $I_m = I/10$ emu or $I_s = (c/10)I$ esu.

Problem 17. Show that the product $\mu_0 K_0$ has the dimensions of the reciprocal of the square of a velocity.

CONVERSION TABLE FOR UNITS

Quantity	Name of practical units	Measure in practical units	Measure in electromagnetic units	Measure in electrostatic units
Charge	Coulomb	q	$q_m = q/10$	$q_s = (c/10)q$
Current	Ampere	I	$I_m = I/10$	$I_s = (c/10)I$
Electric field	$\frac{\text{Volt}}{\text{Meter}}$	E	$E_m = 10^6 E$	$E_s = \frac{1}{3} \times 10^{-4} E$
Potential	Volt	V	$V_m = 10^8 V$	$V_s = \frac{1}{3} \times 10^{-2} V$
Polarization	$\frac{\text{Coulomb}}{\text{Meter}^2}$	P	$P_m = P/10^5$	$P_s = 3 \times 10^5 P$
Displacement	$\frac{\text{Coulomb}}{\text{Meter}^2}$	D	$D_m = 10^{-5} D$	$D_s = 3 \times 10^5 D$
Conductivity	$\frac{\text{Mho}}{\text{Meter}}$	σ	$\sigma_m = 10^{-11}\sigma$	$\sigma_s = 9 \times 10^9 \sigma$
Resistance	Ohm	R	$R_m = 10^9 R$	$R_s = \frac{1}{9} \times 10^{-11} R$
Capacitance	Farad	C	$C_m = 10^{-9} C$	$C_5 = 9 \times 10^{11} C$
Magnetic Field	$\frac{\text{Ampere}}{\text{Meter}}$	H	$H_m = 10^{-3} H$	$H_s = 3 \times 10^7 H$
Induction	$\frac{\text{Weber}}{\text{Meter}^2}$	B	$B_m = 10^4 B$	$B_s = \frac{1}{3} \times 10^{-6} B$
Magnetization	$\frac{\text{Ampere}}{\text{Meter}}$	M	$M_m = 10^{-3} M$	$M_s = 3 \times 10^7 M$

9. Maxwell's Equations in Moving Systems.—The discussion in Sec. 4 showed that the electric field at a point is not an absolute property of the point but depends upon the system of axes with reference to which it is observed. The question as to whether a magnetic field has a similar dependence cannot be answered in so simple a fashion, since the law of force ordinarily used contains no terms that indicate a force on a current moving in an electric field. Nevertheless, if it is desired that Maxwell's equations shall be valid in all systems of coordinates, it is neces-

sary that the magnetic field also shall differ from system to system. Such an assumption is usually made, and the necessary transformation equations are

$$\mathbf{E}' = \mathbf{E} + \mathbf{v} \times \mathbf{B} \qquad \mathbf{B}' = \mathbf{B} - \mu\kappa\mathbf{v} \times \mathbf{E} \tag{15-31}$$

It is important to emphasize the fact that these transformations are adequate only when the relative velocity of the coordinate axes is very much less than the velocity of light but that in this limiting case Maxwell's equations will be approximately valid with reference to all systems. The approximation depends upon $v\sqrt{\mu\kappa}$ and also upon the rate of change with the time of the fields. If these quantities are small enough, the equations are approximately invariant to the transformation. The transformation of the fields that makes the equations exactly valid for all velocities will be taken up in the study of the theory of relativity, where the invariance of these equations will be made the basis of that theory.

Consider equation (IV) in the S' system.

$$\mathbf{curl}'\ \mathbf{H}' = \mathbf{i}' + \frac{\partial \mathbf{D}'}{\partial t'} \tag{IV'}$$

The designation [′] refers to the S' system. In this equation it is important to remember that not only are the field quantities and the current measured with reference to S' but they are expressed in terms of (x',y',z',t') and the derivatives are taken with reference to these coordinates. By means of the transformation equations (15-14), it can be shown that

$$\mathbf{curl}'\ \mathbf{H}' = \mathbf{i}' + \left(v_x \frac{\partial \mathbf{D}'}{\partial x} + v_y \frac{\partial \mathbf{D}'}{\partial y} + v_z \frac{\partial \mathbf{D}'}{\partial z} + \frac{\partial \mathbf{D}'}{\partial t}\right) \tag{15-31a}$$

The last term is due to the fact that $\partial\mathbf{D}'/\partial t'$ implies that the (x',y',z') are held constant, while $\partial\mathbf{D}'/\partial t$ implies that the (x,y,z) are held constant and that $\mathbf{D}'$ is now expressed in terms of (x,y,z,t), even though it is the field measured in S'. The current density in S' is related to that in S by

$$\mathbf{i}' = \mathbf{i} - \rho\mathbf{v} \tag{15-31b}$$

and if the transformation (15-31) is used, (15-31a) becomes

$$\mathbf{curl}\ \mathbf{H} - \kappa\ \mathbf{curl}\ (\mathbf{v} \times \mathbf{E}) = \mathbf{i} - \rho\mathbf{v} + (\mathbf{v} \cdot \nabla)\mathbf{D} + \frac{\partial \mathbf{D}}{\partial t} - (\mathbf{v} \cdot \nabla)\kappa(\mathbf{v} \times \mathbf{B}) - \kappa\mathbf{v} \times \frac{\partial \mathbf{B}}{\partial t} \quad (15\text{-}31c)$$

When $\mathbf{curl}\ (\mathbf{v} \times \kappa\mathbf{E})$ is expanded by means of the vector identity (8) on page 244 and use is made of equation (I) in the S system, this becomes

$$\mathbf{curl}\ \mathbf{H} = \mathbf{i} + \frac{\partial \mathbf{D}}{\partial t} - \mu\kappa\left[(\mathbf{v} \cdot \nabla)(\mathbf{v} \times \mathbf{H}) + \mathbf{v} \times \frac{\partial \mathbf{H}}{\partial t}\right] \quad (15\text{-}31d)$$

This is equation (IV) in the S system except for the final term, which is multiplied by $\mu\kappa$. This product is equal to the reciprocal of the square of the velocity of light in the medium with permittivity κ and permeability μ. As long as v is much less than the velocity of light and $\partial\mathbf{H}/\partial t$ is not too large, the whole term can be neglected. To this approximation the transformation in equation (15-31) makes equation (IV) invariant under a transformation to a moving coordinate system.

Problem 18. Show that the first three of Maxwell's equations are approximately invariant to the transformation from S to S'.

10. Electromagnetic Field Energy.—From the field-theory point of view, the field itself, as well as charged particles of matter, can possess energy and momentum; and, in fact, it must possess these quantities in order to provide for their conservation. Since the particles act, not directly on each other, but only through the intervention of the field, a particle may lose energy and momentum that may not appear immediately in another particle. In the meantime, the missing energy or momentum must be attributed to the field. This idea has already been suggested in the expressions for the energy of electrostatic systems and systems of steady currents that have the form of integrals over the whole field.

In the further discussion of the properties of an electromagnetic field it will always be assumed that the dielectric permittivity κ and the magnetic permeability μ are constant in

both space and time and are independent of the fields. This implies that $\mathbf{D} = \kappa\mathbf{E}$, $\mathbf{B} = \mu\mathbf{H}$ and that κ and μ can be factored out in differentiation. This assumption describes the situation in a vacuum and in a simple homogeneous and isotropic dielectric. It excludes, however, substances in which κ or μ changes from point to point, such substances as iron or nickel, and cases in which several different dielectrics are present. When several homogeneous substances are present, the treatment can be extended by the imposition of suitable boundary conditions.

The general expression for the energy of the field can be developed from Maxwell's equations and the force equation. If both sides of equation (15-29) are multiplied by the velocity and the product is integrated over all the volume in which there is any charge or current, the result is the rate at which the field is doing work on the charges. This will then be the rate at which the energy of the charges is increasing and can be equated to the rate at which the energy of the field is decreasing. Since $\rho\mathbf{v}$ is the current density,

$$-\frac{dW}{dt} = \int \mathbf{F}\cdot\mathbf{v}\,d\tau = \int \rho(\mathbf{E}\cdot\mathbf{v} + \mathbf{v}\times\mathbf{B}\cdot\mathbf{v})d\tau$$
$$= \int \left(\mathbf{E}\cdot\operatorname{curl}\mathbf{H} - \mathbf{E}\cdot\frac{\partial\mathbf{D}}{\partial t}\right)d\tau \qquad (15\text{-}32)$$

The volume element is designated by $d\tau$ to avoid confusion with the velocity. This work is done by the electric field only, since the force exerted by the magnetic field is always perpendicular to the velocity. By the use of a vector transformation and the third of Maxwell's equations, this becomes

$$-\frac{dW}{\partial t} = \int \left(\mathbf{H}\cdot\operatorname{curl}\mathbf{E} - \mathbf{E}\cdot\frac{\partial\mathbf{D}}{\partial t} - div\,\mathbf{E}\times\mathbf{H}\right)d\tau$$
$$= -\int \left(\mathbf{H}\cdot\frac{\partial\mathbf{B}}{\partial t} + \mathbf{E}\cdot\frac{\partial\mathbf{D}}{\partial t}\right)d\tau - \int \mathbf{E}\times\mathbf{H}\,\mathbf{ds} \qquad (15\text{-}33)$$

If the volume over which the integration is carried out is so large that the integral of the fields over the surface vanishes, this rate of change of the field energy may be written

$$\frac{dW}{dt} = \frac{1}{2}\frac{d}{dt}\int (\kappa E^2 + \mu H^2)d\tau \qquad (15\text{-}34)$$

It is very natural to maintain the general validity of the conservation of energy by interpreting the integral in equation (15-34) as the energy in the electromagnetic field. A similarly natural, although perhaps less easily justifiable, interpretation is that the integrand is the volume density of the energy. This conception of localized energy has been fruitful, however, and will be retained as long as it leads to no difficulties.

If integration is carried out over a limited volume, the surface integral does not, in general, vanish. Since the whole expression (15-33) represents the work done by the field on the charges and since the volume integral is to be interpreted as the rate of change of the energy of the field, it is natural to interpret the surface integral as the rate of flow of energy through the surface. The vector $\mathbf{S}$ is called *Poynting's vector*, where

$$\mathbf{S} = \mathbf{E} \times \mathbf{H} \tag{15-35}$$

The interpretation of the surface integral of $\mathbf{S}$ as the flow of energy through the surface leads naturally to the interpretation of $\mathbf{S}$ itself as the density of energy flow.

Problem 19. A wire, which may be treated as approximately straight, carries a steady current I from a battery. Compute Poynting's vector at the surface of the wire, and interpret the result in terms of the energy flow.

Problem 20. A circular loop of wire with resistance $\mathfrak{R}$ is placed in a uniform magnetic field. If the field is changed, find the flow of energy into the wire.

Problem 21. Find Poynting's vector around a uniformly charged sphere placed in a uniform magnetic field.

11. Electromagnetic Field Momentum.—The momentum of the electromagnetic field can be evaluated by the use of arguments similar to those employed for the energy. The rate of increase of momentum of the matter carrying charges and currents will be equal to the volume integral of the force per unit volume. This can then be equated to the rate of decrease of the momentum of the field in order to provide conservation of the total momentum.

$$-\frac{d\mathbf{M}}{dt} = \int \mathbf{F}\, d\tau = \int \rho\mathbf{E}\, d\tau + \int \mathbf{i} \times \mathbf{B}\, d\tau$$
$$= \int \mathbf{E}\, div\, \mathbf{D}\, d\tau + \int (\mathbf{curl}\, \mathbf{H}) \times \mathbf{B}\, d\tau$$
$$- \int \frac{\partial \mathbf{D}}{\partial t} \times \mathbf{B}\, d\tau \quad (15\text{-}36)$$

where $\mathbf{M}$ is the momentum of the field. It follows then, since $\mu\, div\, \mathbf{H} = 0$, that

$$\frac{d\mathbf{M}}{dt} = \frac{d}{dt}\int \mu\kappa\mathbf{S}\, d\tau - \kappa \int (\mathbf{E}\, div\, \mathbf{E} - \mathbf{E} \times \mathbf{curl}\, \mathbf{E}) d\tau$$
$$- \mu \int (\mathbf{H}\, div\, \mathbf{H} - \mathbf{H} \times \mathbf{curl}\, \mathbf{H}) d\tau \quad (15\text{-}37)$$

The last two terms in equation (15-37) can be transformed into surface integrals since they are the divergences of tensors.

The divergence of a symmetric tensor is a vector, which can be defined by

$$(div\, T)_i = \sum_j \frac{\partial T_{ij}}{\partial x_j} \quad (15\text{-}38)$$

By reasoning similar to that used in establishing Gauss's theorem it can be shown that

$$\int_V (div\, T)_i\, d\tau = \sum_j \int_S T_{ij}\, ds_j \quad (15\text{-}38a)$$

where ds_j is the component of the surface perpendicular to the x_j axis.

Let

$$T_{ij}^{(m)} = (H_i H_j - \tfrac{1}{2}H^2\delta_{ij}) \quad (15\text{-}39)$$

and

$$T_{ij}^{(e)} = (E_i E_j - \tfrac{1}{2}E^2\delta_{ij}) \quad (15\text{-}39a)$$

It follows directly from the definitions that

$$div\, T^{(m)} = \mathbf{H}\, div\, \mathbf{H} - \mathbf{H} \times \mathbf{curl}\, \mathbf{H} \quad (15\text{-}39b)$$

and

$$div\, T^{(e)} = \mathbf{E}\, div\, \mathbf{E} - \mathbf{E} \times \mathbf{curl}\, \mathbf{E} \quad (15\text{-}39c)$$

Then

$$\frac{d\mathbf{M}}{dt} = \frac{d}{dt}\int \mu\kappa\mathbf{S}\, d\tau - \int [\kappa T^{(e)} + \mu T^{(m)}]_{ij} ds_j \quad (15\text{-}40)$$

The tensor $\kappa T^{(e)} + \mu T^{(m)}$ is called the *Maxwell stress tensor*. It gives the electromagnetic stress across the surface **ds** as a linear vector function of **ds**. Equation (15-40) then can be interpreted as showing that $\mu\kappa\mathbf{S}$ is the volume density of the momentum in the field and that the rate of change of the field momentum plus the momentum of the charges is equal to the total vector stress on the surface of the portion of the field under consideration.

Problem 22. Evaluate the divergence of the Maxwell stress tensor, and show that equation (15-40) follows from (15-37).

12. General Electromagnetic Potentials.—In the treatment of electrostatics and magnetostatics it was convenient to introduce potentials from which the fields could be derived by differentiation. These potentials were expressed in terms of the charges and the currents by the use of the laws of force between charges or between current elements. Similarly in the general case it is desirable to introduce a scalar potential and a vector potential. These will represent forms in which all solutions of Maxwell's equations can be put, and they will be expressable in terms of the charges and the currents. Such potentials will be similar to the potentials previously used, but they will not in general be the same. Only in static situations will they reduce to the forms already used.

Let Φ be the scalar potential and **A** the vector potential. Then let the fields be derived from the potentials by means of the equations

$$\mathbf{B} = \operatorname{curl} \mathbf{A} \qquad \text{and} \qquad \mathbf{E} = -\operatorname{grad} \Phi - \frac{\partial \mathbf{A}}{\partial t} \tag{15-41}$$

Maxwell's equations (II) and (III) will be identically satisfied by fields derived from any potentials in this way. Equations (15-41) may be considered as general solutions of these two of Maxwell's equations in terms of the arbitrary functions Φ and **A**.

Problem 23. Show that Maxwell's equations (II) and (III) are identically satisfied by any fields that are derived from potentials by means of equations (15-41).

If the potentials are given, equations (15-41) serve to determine the fields uniquely; but if the fields are given, there are many potentials equally satisfactory for describing them. Therefore it is possible to adopt an arbitrary relationship between Φ and $\mathbf{A}$ in order to remove this lack of definiteness. This can be done in a number of ways. To agree with the form of the vector potential used in the chapter on magnetostatics it is necessary to adopt the equation

$$div\ \mathbf{A} = 0 \tag{15-42a}$$

With this condition, equations (15-41) substituted into Maxwell's equations lead to

$$\nabla^2\Phi = -\frac{\rho}{\kappa}$$

$$\nabla^2\mathbf{A} - \mu\kappa\frac{\partial^2\mathbf{A}}{\partial t^2} = -\mu\mathbf{i} + \mu\kappa\ \mathbf{grad}\frac{\partial\Phi}{\partial t} \tag{15-43a}$$

If the potentials Φ and $\mathbf{A}$ satisfy these two equations and if μ and κ are everywhere the same, the fields derived from them by (15-41) will satisfy Maxwell's equations.

The above equations are unsatisfactory for some purposes because the potentials Φ and $\mathbf{A}$ have not been separated. It is possible to separate them if another condition is used instead of (15-42*a*). This is called the *Lorentz condition* and is

$$div\ \mathbf{A} = -\mu\kappa\frac{\partial\Phi}{\partial t} \tag{15-42b}$$

With this condition the equations for the potentials are

$$\left.\begin{aligned}\nabla^2\Phi - \mu\kappa\frac{\partial^2\Phi}{\partial t^2} &= -\frac{\rho}{\kappa}\\ \nabla^2\mathbf{A} - \mu\kappa\frac{\partial^2\mathbf{A}}{\partial t^2} &= -\mu\mathbf{i}\end{aligned}\right\} \tag{15-43b}$$

These equations are called *wave equations* because they represent the propagation of a disturbance with the velocity $1/\sqrt{\mu\kappa}$. General solutions for equations (15-43*b*) can be written in the form of integrals similar to those for the static cases. The difference lies in the fact that the charges and currents effective in

producing the potentials are not those existing at the various points at the time for which the potential is to be computed but are those which existed at such an earlier time that the effect, propagated with the velocity $1/\sqrt{\mu\kappa}$, has just reached the point at which the potential is desired.

Problem 24. Show that, if Φ and $\mathbf{A}$ describe an electromagnetic field, the potentials $\Phi' = \Phi - (\partial f/\partial t)$ and $\mathbf{A}' = \mathbf{A} + \mathbf{grad}\ f$ also describe the same field. f is any scalar function of the coordinates and the time. This transformation of the potentials is called a *gauge transformation.*

Problem 25. Show that, if equations (15-42*a*) and (15-43*a*) are satisfied, the fields will satisfy Maxwell's equations.

Problem 26. Show that, if equations (15-42*b*) and (15-43*b*) are satisfied, the fields will satisfy Maxwell's equations.

Problem 27. Show that the potentials of equation (15-43*a*) can be derived from those of (15-43*b*) by means of a gauge transformation, and find the differential equation for the function f.

13. Electromagnetic Waves in Homogeneous Uncharged Dielectrics.—In a homogeneous uncharged dielectric, Maxwell's equations lead to wave equations for the fields.

$$\nabla^2\mathbf{E} - \mu\kappa\frac{\partial^2\mathbf{E}}{\partial t^2} = 0 \qquad \nabla^2\mathbf{H} - \mu\kappa\frac{\partial^2\mathbf{H}}{\partial t^2} = 0 \tag{15-44}$$

These two second-order equations are more general than the first-order equations from which they are derived. Their general solution is not itself a solution of Maxwell's equations, but it must be limited by additional conditions. Since they are partial differential equations, they have a wide variety of particular solutions. One kind of solution is of the form

$$\mathbf{E} = \mathbf{f}(\alpha x + \beta y + \gamma z - vt) \tag{15-45}$$

where $v^2 = 1/\mu\kappa$. This represents a wave in the electric field. The form of the wave is given by the arbitrary vector function f, the direction of propagation is given by α, β, γ, and the velocity of propagation is v. According to equations (15-44) the magnetic field could contain another similar, but independent wave. However, the original Maxwell's equations require a connection

between the electric and magnetic fields so that the two solutions cannot be independent.

Problem 28. Derive equations (15-44).

Problem 29. Show that (15-45) is a solution of (15-44).

Problem 30. Show that, in the absence of charges, the first of Maxwell's equations requires the electric field in a plane electromagnetic wave of the form (15-45) to be perpendicular to the direction of propagation.

Problem 31. Show that Maxwell's equations require the electric and the magnetic fields in a plane wave to be perpendicular to each other and to the direction of propagation. This assumes that no constant fields are present.

Problem 32. If a plane electromagnetic wave is sinusoidal, has its electric vector always along the x axis, and is propagated in the z direction, the expression for the electric vector can be written in the form

$$\mathbf{E} = \mathbf{i}E_x = \mathbf{i}E_0 \sin a(z - vt) \qquad (15\text{-}46,$$

Interpret the constants in this equation in terms of the amplitude) wave length, and frequency of the wave.

Problem 33. Find the magnetic wave that accompanies the electric wave of equation (15-46).

Problem 34. Find the energy density in the wave of the above two problems.

Problem 35. Find Poynting's vector and the momentum density in the wave of Probs. 32 and 33.

Problem 36. Show that the wave equation (15-44) is satisfied by

$$\mathbf{E} = \mathbf{i}E_x = \frac{\mathbf{i}f(r - vt)}{r} \qquad (15\text{-}47)$$

14. Lorentz's Form of Maxwell's Equations.—According to the electrical theory of matter, the dielectric permittivity and the permeability are quantities that describe average properties of material substances. The idea is that, if the materials were examined in detail, only the motions of charged particles would have to be considered. The development of this point of view constitutes the electron theory, much of which is associated with the name of H. A. Lorentz. When Maxwell's equations are written for use from this point of view, the only field quantities that can appear are **E** and **B** and the current can be replaced

by the motion of the charges. Hence the equations become

$$\left.\begin{aligned} &(\text{I}a)\ div\ \mathbf{E} = \frac{\rho}{\kappa_0} \qquad (\text{II}a)\ div\ \mathbf{B} = 0 \\ &(\text{III}a)\ \mathbf{curl}\ \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \\ &(\text{IV}a)\ \mathbf{curl}\ \mathbf{B} = \mu_0\rho\mathbf{v} + \mu_0\kappa_0\frac{\partial \mathbf{E}}{\partial t} \end{aligned}\right\} \tag{15-48}$$

The quantity $\mathbf{v}$ is the vector velocity of the charge density ρ.

Although the equations when written in this form constitute the starting point for the classical electron theory, the electrons really have no place in them. It is the charge density and not the number of electrons that appears. This fact implies that the equations are to be used to describe the structure of the electrons themselves, and many attempts have been made to find a solution that would represent a small amount of charge permanently held together by electromagnetic forces. All such attempts have failed. The classical electron theory has never been able to account satisfactorily for the fact that electrons exist, and this remains one of the outstanding problems of theoretical physics.

In spite of the incompleteness of the theory, it is possible to use equations (15-48) to describe the fields of electrons at distances from them large compared with their dimensions. Because the electrons are very small, this makes the equations of considerable use.

15. Radiation from an Oscillating Dipole.—The light waves sent out by an excited atom, as well as the long waves emitted by a radio antenna, can be closely approximated by the radiation of an oscillating dipole. The dipole may be pictured as made up of two spherical conductors, separated by a distance greater than their diameters, and connected by a wire. It will be assumed that energy is supplied to the system to cause the charge to oscillate between the spheres in a prescribed sinusoidal manner. The system will then have an electric moment determined by the charges on the spheres, and this electric moment will vary sinusoidally with the time. The current in the connecting wire, which will be proportional to the rate of change

of the charge on the spheres, will also vary sinusoidally. The vector potential around the system will always have the direction of the current in the connecting wire. This direction will be taken as the axis of a system of polar coordinates and as the z axis of a Cartesian system. The origin of the coordinates will be at the center of the dipole. The values of the field quantities and the potentials will be determined only at distances from the origin that are much greater than the length of the dipole h. It must also be assumed that the frequency of the sinusoidal variation is low enough so that the changes which take place in the time h/v can be neglected.

Under these conditions the vector potential at distances from the origin large compared with h can be written

$$\mathbf{A} = A_z\mathbf{k} = -\frac{\mathbf{k}f'(r - vt)}{r} \tag{15-49}$$

In this expression $f'(r - vt)$ is the derivative, with respect to its argument, of a function $f(r - vt)$ to be evaluated later. It can be shown by substitution that this expression for the vector potential satisfies equation (15-43*b*), and it will be shown later in what way it satisfies the boundary conditions. The combination of equation (15-49) with equation (15-42*b*) then permits the determination of the scalar potential, with the exception of a constant, which is unimportant here. The result is

$$\Phi = \frac{vz}{r^3} f(r - vt) - \frac{vz}{r^2} f'(r - vt) \tag{15-50}$$

In the region close to the dipole, although still for values of r much larger than h, the presence of r in the argument of f can be neglected, and the term with the higher power of r in the denominator will be the larger. That such is the case is due to the restriction imposed on the rate of change of the current. Hence the principal term in the scalar potential, in this region, is

$$\Phi = \frac{v}{r^2} f(-vt) \cos\theta \tag{15-51}$$

This expression is the ordinary electrostatic potential of a dipole of magnitude $4\pi\kappa_0 vf(-vt)$ parallel to the z axis. Since, in the

immediate neighborhood of the dipole, the time necessary for the propagation of the effects can be neglected, the result indicates that the potential assumed is correct to this extent.

In the region farther from the dipole, where the current has changed considerably in the time r/v, only that term in the potential need be retained which has the lower power of r in the denominator. In this region,

$$\Phi = -\frac{v}{r} f'(r - vt) \cos \theta \qquad \mathbf{A} = -\mathbf{k} \frac{f'(r - vt)}{r} \tag{15-52}$$

This zone is called the *wave zone,* because the predominating term in the potentials represents a wave traveling outward from the origin.

Problem 37. Show that in the wave zone the electric and magnetic fields are perpendicular to each other and to the radius vector. Use polar coordinates.

Problem 38. Compute Poynting's vector in the wvae zone, and show that the total energy which passes outward through a spherical shell per unit time is

$$U = \frac{8\pi}{3} \sqrt{\frac{\kappa}{\mu}} \left[\frac{d^2 f(-vt)}{dt^2}\right]^2_{t-(r/v)} \tag{15-53}$$

For this result it is not required that the current in the dipole vary sinusoidally, but the subscript indicates that the second derivative is to be taken at the time r/v before the time considered. This is the time necessary for the energy to travel from the dipole to the spherical shell.

Problem 39. Show that if the current in the dipole varies sinusoidally with the time the average rate of emission of energy is proportional to the fourth power of the frequency.

References

ABRAHAM-BECKER: "Classical Electricity and Magnetism," Part III, Blackie & Son, Ltd., Glasgow.

PAGE, L., and N. I. ADAMS: "Principles of Electricity," Chap. 16, D. Van Nostrand Company, Inc., New York.

PAGE, L.: "Introduction to Theoretical Physics," Chap. 12, D. Van Nostrand Company, Inc., New York.

MASON, MAX, and WARREN WEAVER: "The Electromagnetic Field," Chap. 4, University of Chicago Press, Chicago.

SMYTHE, WILLIAM R.: "Static and Dynamic Electricity," McGraw-Hill Book Company, Inc., New York.

CHAPTER XVI

THE RESTRICTED THEORY OF RELATIVITY

The notation of vector analysis provides a method by which many of the equations of physics can be written in a form independent of the orientation and position of the coordinate system used to locate the bodies. This notation does not, however, remove the distinction between systems of coordinates that are moving relative to each other. Some use was made of moving systems in the study of solid bodies, but the vectors were always, in the last analysis, referred to coordinates at rest with respect to the observer and for which Newton's laws were postulated. In the study of electromagnetic fields some attention was given to moving systems of coordinates, but it was specified that only velocities small compared with the velocity of light were to be considered. The theory of relativity consists in the study of consequences of the assumption that all coordinate systems are equally good for the description of physical phenomena. The restricted theory treats only those systems of coordinates that move relative to each other with constant velocity. Other cases are included in the general theory of relativity.

1. Invariance of Newton's Laws under Galilean Transformations.—Newton's equations [equations (2-2)] are unchanged when the coordinates and the time are subjected to a Galilean transformation. A Galilean transformation is given by the equations

$$x' = x - vt \qquad y' = y \qquad z' = z \qquad t' = t \tag{16-1}$$

According to the elementary ideas of space and time, this represents a transformation to a system of coordinates S' moving with the velocity v in the positive x direction with respect to the system S. This is a specialized form of the most general Galilean transformation. It is obtained by assuming that, when $t = t' = 0$, the origins of the two systems of coordinates

coincide. It is also assumed that the motion is in the direction of the x axis. These restrictions do not really limit the generality of the results, and therefore they will always be used here and in the later treatment of the Lorentz transformation.

The invariance of Newton's laws is descriptive of the fact that mechanical processes should appear the same to all observers who move relative to one another with uniform velocity. This certainly does not mean that velocities and positions are the same for all observers, or even that such quantities as energy and momentum are the same; but it does mean that the relationships suitable for the fundamental laws cannot depend on which particular coordinate system is used. In speaking of the transformation from one coordinate system to another, it is implied that there is an observer in each coordinate system who observes the physical events in terms of his coordinates and time. The transformation equations then express the relationship between the observations, by different observers, of the same events.

The transformation of Newton's laws of motion involves the transformation of second derivatives with respect to the time. It follows directly from equation (16-1) that $d^2x'/dt'^2 = d^2x/dt^2$ and that the forces as functions of (x,y,z,t) can be expressed in terms of (x',y',z',t'). Hence the laws have the same form in the system S' and S.

Problem 1. Find the way in which momentum and kinetic energy transform under a Galilean transformation.

2. The Postulates of Relativity.—The first postulate of relativity is an extended form of an old idea. Because of the invariance of Newton's equations with respect to a Galilean transformation, it is impossible, by means of any mechanical experiment, to detect any such thing as an "absolute velocity." An absolute velocity would be a velocity relative to some system of coordinates which is a "preferred" system, in that the laws of mechanics might take some especially simple form when referred to it. Since no such thing as absolute velocity can be detected, it cannot be used in the formulation of the theory. Hence, in the realm of Newtonian mechanics, all velocity is relative.

The first postulate of relativity extends this idea to all physics. This postulate is based on the fact that no experiment has ever been devised which detected any motion of the earth through an "ether," although many were tried. The postulate, then, can be stated as follows:

1. All systems of coordinates are equally suitable for the description of physical phenomena.

This postulate literally refers to all systems of coordinates, although the treatment here is restricted to those that are moving relative to each other with uniform velocity. The more general treatment is called the *general theory of relativity.*

There are a number of equivalent forms in which this first postulate can be stated. It is equivalent to the statement that all velocities are relative and that it is impossible by any means whatever to detect a motion through an ether. In spite of this principle, Maxwell's equations seem to indicate such a possibility. Maxwell's equations contain first derivatives and are not invariant to the Galilean transformation. This could mean that the first postulate of relativity is not true, or it might mean that Maxwell's equations are not exact but are only approximations to the correct laws. Einstein, however, made the startling assumptions that the first postulate is true and at the same time Maxwell's equations are exact. If both these things are assumed, the only escape from the apparent inconsistency is that the Galilean transformation is not the correct way in which to relate observations made with reference to different coordinate systems.

Instead of postulating the invariance of Maxwell's equations, it is sufficient to make the second postulate as follows:

2. The velocity of light is independent of the relative motion of the source and the observer.

It is this second postulate that leads to the unfamiliar results of the Einstein theory of relativity. This postulate could be stated with reference to the Maxwell equations. It could be postulated that these equations must be invariant to a transformation of coordinates. However, the above statement is simpler and can be shown to amount to the same thing. This postulate then implies that the Maxwell equations have the

correct form for the representation of electromagnetic phenomena but that Newton's equations do not have the correct form for the exact representation of mechanical phenomena.

With these two postulates as the basis of the theory of relativity, it is first necessary to find the transformation of coordinates that will satisfy the second postulate and then to find such laws of mechanics as will satisfy the first postulate with this new form of transformation. The applicability of the two postulates to natural phenomena can then be appraised by the correctness of the results deduced from them.

3. The Lorentz Transformation.—It is easy to see that the Galilean transformation does not satisfy the second postulate of relativity; therfore, if this postulate is to be adopted, it is necessary to find another transformation. The transformation that has been adopted is called the *Lorentz transformation* because of the contributions made by H. A. Lorentz to its development.

In a system of coordinates in which a source of light is at rest at the origin, the equation for the propagation of a pulse of light is

$$x^2 + y^2 + z^2 - c^2t^2 = 0 \tag{16-2}$$

In another system of coordinates moving relative to the first with a uniform velocity, the second postulate requires that the propagation of the same pulse of light be described by

$$x'^2 + y'^2 + z'^2 - c^2t'^2 = 0 \tag{16-2a}$$

This assumes, of course, that the origins of the two systems coincide at the time at which the light pulse starts. Such a restriction does not at all limit the generality of the result. The requirement of the second postulate is that the equations of transformation connecting the description of the event in the S' system with the description of the same event in the S system shall leave the form of equation (16-2) invariant.

The left-hand side of equation (16-2) may be regarded as the length of a four-dimensional vector in a space whose coordinates are x, y, z, and ict. The transformations that will leave (16-2) invariant are then the orthogonal transformations in four

dimensions. With this kind of representation, a transformation to a moving system of coordinates consists in a rotation of the four-dimensional coordinates. Care must be taken in the use of this system to remember that one of the coordinates is imaginary, but for many purposes this makes no difference. As mentioned before, the analysis is simplified, and no generality is lost, if it is assumed that the relative motion of the two three-dimensional coordinate systems is parallel to their x axes. The corresponding rotation of the four-dimensional system is around the y and the z axes, and thus these are unchanged. The matrix of the transformation then takes the form

$$\begin{pmatrix} \gamma_{11'} & 0 & 0 & \gamma_{41'} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \gamma_{14'} & 0 & 0 & \gamma_{44'} \end{pmatrix} \tag{16-3}$$

The coordinates x_1, x_2, x_3, x_4 are the space and time coordinates x, y, z, and ict, respectively.

The four coefficients left undetermined in this transformation can be reduced to a single coefficient β, by the use of the conditions for orthogonality. Let $\gamma_{11'} = \alpha$, and let $\gamma_{41'} = i\alpha\beta$. Then the orthogonality conditions and the requirement that the determinant of the transformation shall be equal to unity give $\gamma_{14'} = \mp i\alpha\beta$, $\gamma_{44'} = \pm\alpha$, and $\alpha = 1/\sqrt{1 - \beta^2}$. In order that the time shall not be reversed, it is necessary to take the upper sign all the way through. If, then, the transformation is written in terms of the original space and time coordinates, the equations are

$$\left.\begin{aligned} x' &= \frac{x - \beta ct}{\sqrt{1 - \beta^2}} & y' &= y \\ t' &= \frac{t - \beta x/c}{\sqrt{1 - \beta^2}} & z' &= z \end{aligned}\right\} \tag{16-3a}$$

This is the usual form of the Lorentz transformation.

For the zero value of x' in equations (16-3a), $x = \beta ct$. This means that the system S is moving past the system S' with the speed βc in the negative x direction. Similarly, then, the system S' is moving past the system S with the same speed in

the positive x direction. This determines β in terms of the relative velocity of the systems.

Problem 2. Work out the details of the process by which (16-3a) is shown to be the necessary form for an orthogonal transformation in the coordinates x_1, x_2, x_3, x_4.

Problem 3. Invert the Lorentz transformation as given in (16-3a) to get the expression for the unprimed quantities in terms of the primed.

Problem 4. Show by direct application of the Lorentz transformation that the expression (16-2) is left invariant.

Problem 5. Show that, as $c \to \infty$, the Lorentz transformation approaches the Galilean.

The treatment just given has shown only that the second postulate of relativity is satisfied if the space and time coordinates transform according to the Lorentz transformation. It can be shown, however, that the Lorentz transformation is a consequence of the postulate; and since there is a considerable amount of experimental evidence in favor of the postulate, there is good reason for concluding that the Lorentz transformation is the proper one by means of which to connect observations made in different coordinate systems.

It is important to understand the sequence of the argument. The quantities x, y, z, t are the space and time coordinates of a definite physical event as it is described in one system of coordinates. The quantities x', y', z', t' are the corresponding coordinates of the same physical event when it is described in another system of coordinates. The first postulate states that one system is as good as another for the description of physical phenomena, and hence a correctly formulated physical law must have the same form in both systems. It is then assumed in the second postulate that the spherical propagation of light with constant velocity is one such correctly stated law. Because of this, it must be the same in all coordinate systems so that the Lorentz transformation is required as the correct way to correlate observations made with reference to one system with those made with reference to another. Physical laws that change their essential form when this transformation is made are then considered as incorrectly formulated.

4. Transformation of Maxwell's Equations.—If any equation representing a physical law can be written as an equation between vectors in the kind of four-dimensional space-time described above, it will be invariant to the Lorentz transformation and thus will satisfy the postulates of relativity. Since all electromagnetic phenomena are described by Maxwell's equations, it is necessary to investigate the possibility of writing these in a four-dimensional vector form.

The electromagnetic potentials introduced in the previous chapter form a four-dimensional vector. We shall be concerned here only with the Lorentz form of the equations; thus the permittivity and the magnetic permeability can be given the values for a vacuum, and $c^2 = 1/\mu_0\kappa_0$. It is then possible to define a four-dimensional potential Φ_k and a four-dimensional current P_k as shown in the following table:

k	1	2	3	4
x_k	x	y	z	ict
Φ_k	A_x	A_y	A_z	$i\Phi/c$
P_k	$\mu_0\rho v_x$	$\mu_0\rho v_y$	$\mu_0\rho v_z$	$i\sqrt{\frac{\mu_0}{\kappa_0}}\rho$

In this table the A's are the components of the magnetic vector potential, Φ is the scalar potential, and the v's are the components of the velocity of the charge density ρ. All the quantities are functions of the four coordinates, so that Φ_k and P_k are two vector fields in four dimensions.

In the four-dimensional vector analysis there are operations analogous to those in three dimensions, although there are some important differences. A four-dimensional gradient can be defined by

$$\textbf{Grad } U = \frac{\partial U}{\partial x_k}\,\mathbf{e}_k \tag{16-5}$$

In this equation the summation convention implies a sum of four terms on the right-hand side, and the $\mathbf{e}_k$'s are the four unit

vectors in the directions of the four coordinate axes. The symbolic operator ∇ is written $\square$ in four dimensions, and the divergence can be written

$$Div\ \mathbf{A} = \square \cdot \mathbf{A} = \frac{\partial A_k}{\partial x_k} \tag{16-6}$$

The operator analogous to the Laplacian is

$$Div\ \mathbf{Grad}\ U = \square^2 U = \sum_k \frac{\partial^2 U}{\partial x_k{}^2}$$
$$= \frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} + \frac{\partial^2 U}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2 U}{\partial t^2} \tag{16-7}$$

With this notation, the equations obeyed by the electromagnetic potential can be written in very concise form.

$$\square^2 \Phi_k = -P_k \qquad \text{with } Div\ \Phi = 0 \tag{16-8}$$

Since this is an equation between vectors, it will be invariant under orthogonal transformations and thus will be invariant under the Lorentz transformation.

The electromagnetic fields can be obtained from the four-dimensional potential by an operation analogous to the curl. Here, however, the presence of four coordinates makes evident a distinction concealed in the three-dimensional analysis, although even in three dimensions the curl is an operation connected with two axes rather than with one. The x component of the curl is obtained by differentiation with respect to the y and the z-axes. However, since there is only one axis perpendicular to these two, the result of the differentiation can be associated with this axis in the same way that an area is represented by a vector perpendicular to it. In four dimensions such an association is not possible, and the curl must be treated as a tensor of the second rank. A tensor of the second rank in four dimensions has formally 16 components. The Curl, however, is an antisymmetric tensor in which the diagonal components are zero and the others are oppositely equal in pairs, so that there are only six independent components. For this reason it is sometimes called a *six-vector*. If the field tensor is designated by f_{hk}, it is

defined by

$$f_{kh} = \frac{\partial \Phi_h}{\partial x_k} - \frac{\partial \Phi_k}{\partial x_h} \tag{16-9}$$

In terms of this field tensor Maxwell's equations can be written very compactly.

$$\frac{\partial f_{hk}}{\partial x_k} = P_h \qquad \text{and} \qquad \frac{\partial f_{jk}}{\partial x_h} = 0 \tag{16-10}$$

The first of these equations is written with the ordinary summation convention so that there are four separate equations, each of which contains four terms on the left-hand side. The second equation contains three different subscripts, representing three out of the four possible dimensions. These are to be taken in the cyclical order h, j, k, and the sum is to be taken of the terms in which h takes the three possible values in succession. In case the omitted index is 2, the equation will read

$$\frac{\partial f_{34}}{\partial x_1} + \frac{\partial f_{41}}{\partial x_3} + \frac{\partial f_{13}}{\partial x_4} = 0 \tag{16-10a}$$

Problem 6. Show that equation (16-8) is the equation for the electromagnetic potentials.

Problem 7. Show that the field tensor is antisymmetrical.

Problem 8. Show that the field tensor f_{hk} is associated with the ordinary field quantities as follows:

$$\begin{pmatrix} f_{11} & f_{12} & f_{13} & f_{14} \\ f_{21} & f_{22} & f_{23} & f_{24} \\ f_{31} & f_{32} & f_{33} & f_{34} \\ f_{41} & f_{42} & f_{43} & f_{44} \end{pmatrix} \equiv \begin{pmatrix} 0 & B_z & -B_y & -iE_x/c \\ -B_z & 0 & B_x & -iE_y/c \\ B_y & -B_x & 0 & -iE_z/c \\ iE_x/c & iE_y/c & iE_z/c & 0 \end{pmatrix} \tag{16-11}$$

Problem 9. Show that equations (16-10) are Maxwell's equations for vacuum.

The four-dimensional form in which Maxwell's equations can be written shows very clearly their invariance under the Lorentz transformation, and it also brings out the intimate connection between electric and magnetic phenomena. In the previous chapter it was shown that an electric field with reference to one system of coordinate axes may be part electric and part magnetic when referred to another set of axes. The

previous considerations were restricted to low relative velocities and were approximate. They are now extended to include all velocities. The four-dimensional transformation provides exact invariance for the Maxwell equations and shows that these equations are applicable, in all systems of coordinates, and that pure electric and magnetic fields are merely limiting cases of the general electromagnetic field.

Problem 10. Apply the Lorentz transformation to the four-dimensional electromagnetic potential to determine the equations of transformation for the magnetic vector potential and the scalar electric potential between systems of coordinates moving with respect to each other.

Problem 11. The general formula for the transformation of a tensor of the second rank by an orthogonal transformation is

$$f_{hk} = \gamma_{hg'} f_{g'i'} \gamma_{i'k}$$

With this transformation formula show that the transformation equations for the field are

$$\begin{aligned} E_{x'} &= E_x & B_{x'} &= B_x \\ E_{y'} &= \frac{E_y - \beta c B_z}{\sqrt{1-\beta^2}} & B_{y'} &= \frac{B_y + \beta E_z/c}{\sqrt{1-\beta^2}} \\ E_{z'} &= \frac{E_z + \beta c B_y}{\sqrt{1-\beta^2}} & B_{z'} &= \frac{B_z - \beta E_y/c}{\sqrt{1-\beta^2}} \end{aligned} \tag{16-12}$$

Problem 12. Show that the transformations in (16-12) approach those of Chap. XV when the relative velocity of the coordinate systems is small compared with the velocity of light.

It is possible to use these transformation equations to find the fields due to charges moving with uniform velocities. All that is necessary is to take the fields present when the charges are at rest relative to the observer and then to transform them to a system of coordinates in which the charges appear to be moving. Thus the observer with reference to which a charged particle is moving will see a magnetic field, while an observer who is moving with the charge will see only an electrostatic field.

Problem 13. Find the electric and magnetic fields that surround a charged particle moving with uniform velocity along the x axis.

Problem 14. Find the electric and magnetic fields surrounding a long charged wire moving parallel to its length.

Problem 15. Find the electric and magnetic fields about a long charged wire moving perpendicular to its length.

5. Consequences of the Lorentz Transformation for Mechanics.—The discussion in the preceding section has shown that, if the Lorentz transformation is the proper way in which to connect the coordinates and time measured in one system of coordinates with those measured in another system, electromagnetic phenomena will appear to follow Maxwell's equations in both sets of coordinates. It will then never be possible to tell anything about an "absolute motion" by performing electromagnetic experiments. Since no experiment has ever given any conclusive indication of an absolute motion and since the velocity of light appears experimentally to be independent of the relative motion of the source and the observer, it has been concluded that the Lorentz transformation equations are the correct ones to use. However, these transformation equations do not leave Newton's equations invariant. If the first postulate of relativity is to be maintained for mechanical experiments, it is necessary to conclude that Newton's equations are not exact and require some corrections. Most of this consists in a modification of the kinematics, but there is also a modification of the mechanical concepts of energy and mass.

a. The Relativity of Simultaneity.—The use of the Lorentz transformation implies that events which are simultaneous as seen from one system of coordinates are not necessarily simultaneous as seen from another system, moving relative to the first. This lack of simultaneity appears only when the events do not coincide in space. If two events coincide both in space and in time as seen from one system of coordinates, they will similarly coincide in all others. Relationships of this kind can be determined by imagining a definite physical situation in one set of coordinates and then transforming it to the other.

Problem 16. Show that, if two events are simultaneous in the system S in which they have different space coordinates, they will not in general be simultaneous as seen from the system S' but will be

separated by the time interval

$$\Delta t' = \frac{\beta \, \Delta x/c}{\sqrt{1 - \beta^2}} \tag{16-13}$$

Problem 17. Show that, if two events are separated by the time interval Δt in the S system in which they have the same space coordinates, they will appear from the S' system to be separated by the time interval

$$\Delta t' = \frac{\Delta t}{\sqrt{1 - \beta^2}} \tag{16-14}$$

Problem 18. Show that the above result is reciprocal, *i.e.*, that two events which occur at the same place in the S' system appear to an observer in the S system to be separated by a time interval longer than that measured by an observer in the S' system. Show that this result cannot be obtained by solving equation (16-14) for Δt, because the conditions for the validity of equation (16-14) are not satisfied.

b. The Lorentz Contraction.—One of the best-known effects of the Lorentz transformation is the so-called "Lorentz contraction." This is in some respects an unfortunate name, as it implies that the material of a body contracts when it is set in motion. The viewpoint of relativity, however, is that this apparent contraction is not due to any special property of the body itself but is a general property of space and time.

Consider a measuring stick with its length parallel to the x axis. In the S system, let one end of the stick be at the origin and the other at the point $x = L$. Then consider a system S' that is moving relative to S and to the stick with the velocity v in the x direction. To find the length of the stick as seen from this system, it is necessary to observe the x' coordinates of the two ends of the stick and to observe them at the same time, *i.e.*, the same value of t'. For simplicity let $t' = 0$. Then, although these observations are simultaneous in the S' system, they will not be simultaneous in the S system. In fact, from the equations of transformation the time at which the one end of the stick is observed will be $t = 0$, while the time of observation of the other end of the stick will be $t = \beta L/c$. These are the values of t that must be put into the transformation equations to determine the values of x' that will be observed. The result is

that

$$L' = L\sqrt{1 - \beta^2} \tag{16-15}$$

Problem 19. Derive equation (16-15) in detail.

c. The Transformation of Velocity.—The measurement of a uniform velocity consists in the observation of two events. Each observation consists in the measurement of a coordinate and the corresponding time. If it is observed from the system S that a particle has the coordinate x_1 at the time t_1 and then has the coordinate x_2 at the time t_2, the average velocity of the particle is, by definition, given by

$$V_x = \frac{x_2 - x_1}{t_2 - t_1} \tag{16-16}$$

By the transformation of these four individual coordinates it is possible to obtain the equation for the transformation of velocities.

Problem 20. Carry out the above described analysis, and show that the transformation of a velocity in the x direction is given by

$$V_{x'} = \frac{V_x - \beta c}{1 - (\beta V_x/c)} \tag{16-17}$$

Problem 21. Show that the transformation for a velocity perpendicular to the direction of motion of the coordinate system is given by

$$V_{y'} = V_y(1 - \beta^2)^{\frac{1}{2}} \tag{16-18}$$

Problem 22. Carry through similar considerations for the transformation of a velocity that has both x and y components.

Problem 23. Show from the above transformation equations for velocity that it is impossible, by adding velocities, to get a velocity greater than the velocity of light.

d. The Laws of Motion and the Transformation of Mass.—Although the ordinary Newtonian laws of motion are not invariant under the Lorentz transformation and hence cannot be considered as exact, it is possible to set up very similar laws that do hold under the postulates of relativity. The difference consists essentially in a change in the concepts of mass and

energy. Newton's equations may be taken over in the form that he originally gave them, *i.e.*,

$$\frac{dM_h}{dt} = F_h \tag{16-19}$$

where M_h is the component of the momentum along the h axis and F_h is the corresponding component of force. There can be no question about the validity of this equation when it is regarded as a definition of force, but it is then necessary to consider carefully the definition of the momentum.

Momentum can be defined by Newton's third law. According to this, if two bodies influence each other, the change of momentum of one is equal and opposite to the change of momentum of the other. This statement can be used for the definition of momentum by carrying out an imaginary experiment first proposed by Tolman and Lewis.

If a particle is moving with a velocity very small compared with the velocity of light, the momentum must be given by the ordinary expression $m_0 v$. This is because the Lorentz transformation reduces to the Galilean when the light velocity approaches infinity. The mass is written as m_0 to indicate this restriction to small velocities. Since the velocity can be measured and defined in all systems of coordinates, the expression mv can always be used for the momentum if m is allowed to vary with the velocity. It is then necessary that m approach m_0 as the velocity approaches zero. The way in which m varies with the velocity can be determined from the conceptual experiment of Tolman and Lewis.

Consider two systems of coordinates S and S'. Let them be moving past each other along their coincident x axes with the velocity $v = \beta c$. Let there be a particle whose mass, when at rest, is m_0, and which has the velocity u_y along the y axis in the S system, and let there be an exactly similar particle with the velocity $-u_{y'}$ in the S' system. Each of these particles is moving parallel to the y axis of its own system with a speed small compared with v. Let the two particles be moving in such a way that they collide and each is reversed in its direction of motion as seen from its own system. It is thus assumed that

neither particle acquires in the collision any motion along the x or z axes. If $u_y = u_{y'}$ before the collision, the symmetry of the situation and the first postulate of relativity require that $U_y = U_{y'}$, where these are the component velocities after the collision. Thus far the experiment has been described by giving the motion of each particle in its own system of coordinates. Now consider the whole experiment as seen from the S system. The particle that has no x or z velocity in the S' system will appear to have the velocity components

$$w_y = u_{y'}\sqrt{1 - \beta^2} \qquad \text{and} \qquad W_y = U_{y'}\sqrt{1 - \beta^2} \tag{16-20}$$

before and after the collision. This represents the application of the Lorentz transformation. The application of the postulate, or definition, of the conservation of momentum then leads to the equation

$$\begin{aligned} m_0u_y + m_0U_y &= m'w_y + m'W_y \\ &= m'(u_{y'} + U_{y'})\sqrt{1 - \beta^2} \end{aligned} \tag{16-21}$$

This leads to the conclusion that

$$m' = \frac{m_0}{\sqrt{1 - \beta^2}} \tag{16-22}$$

If the mass is defined to vary with the velocity as indicated in equation (16-22), the momentum can be expressed as mv and there will be conservation of momentum.

In the above treatment the variation of the mass due to the motion of the particles along their y axes has been neglected. This is of course only justified because $u_y \ll v$. However, it can be shown by a more detailed analysis that the result in equation (16-22) still holds when this is taken into account.

With the above definitions of mass and momentum, equation (16-19) gives a definition of force. The law of transformation for force can be determined from the law of transformation of the left-hand side.

e. Equivalence of Mass and Energy.—With the definitions of mass, momentum, and force just indicated, a very remarkable theorem concerning the equivalence of mass and energy can be

established. Equation (16-19) can be written

$$\frac{d}{dt}(m\mathbf{v}) = m\frac{d\mathbf{v}}{dt} + \mathbf{v}\frac{dm}{dt} = \mathbf{F} \tag{16-23}$$

If the increase in the energy of a particle is defined as equal to the amount of work done on it, the principal of the conservation of energy can be maintained. Hence let

$$\begin{aligned}\Delta E = \int \mathbf{F}\cdot\mathbf{dl} &= \int d(m\mathbf{v})\cdot\mathbf{v} = \int m\mathbf{v}\cdot\mathbf{dv} + \int \mathbf{v}\cdot\mathbf{v}\,dm \\ &= \int v^2\,dm + \tfrac{1}{2}\int m d(v^2)\end{aligned} \tag{16-24}$$

By means of equation (16-22) it is possible to express the speed as a function of the mass, so that the integral in (16-24) can be expressed as an integration with reference to the mass. The result is

$$\Delta E = \int \left(v^2 + \frac{m_0{}^2}{m^2}c^2\right) dm = c^2 \int_m^{m+\Delta m} dm = c^2\,\Delta m \tag{16-25}$$

This is the important result that the increase in the energy of a particle is just equal to c^2 times the increase in mass. It suggests that the total mass can be expressed as energy in the equation

$$E = mc^2 \tag{16-26}$$

The derivation just given does not fix the constant of integration and does not require equation (16-26). The early experiments on the mass of electrons at high velocities led to confirmation of equation (16-22) for the change of mass with velocity, but the first conclusive confirmation of equation (16-26) came in the study of nuclear physics, where this equivalence of mass and energy is now one of the fundamental principles.

Problem 24. Consider an electron as a charged sphere inside of which there is no field. By the use of equation (16-26), compute the mass of the electron as a function of the radius of the sphere and the charge on it. The energy of the electron is that of the electric field around it.

Problem 25. In the same way compute the mass of a moving electron. Note that when in motion the sphere is subject to the Lorentz transformation and becomes an ellipsoid of rotation.

Problem 26. Show by application of the Lorentz transformation that the force on a charged particle which is moving with uniform velocity is just equal to that due to the electrostatic field as seen by an observer who is at rest with respect to the particle.

The postulates of relativity have been developed into a very elaborate structure, and the necessity for invariance, under the Lorentz transformation, of any physical law has been generally recognized. The restricted theory has been important in treating the high-speed motions of atomic and nuclear particles. The general theory has been of importance in the treatment of large-scale cosmological problems, where gravitational forces and potentials are of major importance. However, there still remain important problems that are not understood.

References

TOLMAN, R. C.: "Relativity, Thermodynamics, and Cosmology," Oxford University Press, New York.

RIEMANN-WEBER: "Differentialgleichungen der Physik," Part IV (by Sommerfeld, on Electromagnetic Vibrations), Friedrich Vieweg & Sohn, Brunswick.

INDEX

A

Absolute scale of temperature, 204
Absolute value of a complex number, 52
Absorption of energy by an oscillator, 78
Adiabatic process, 197
Air, resistance of, 29
Amperes, 284
Ampère's laws, 289
Analogies, thermodynamic, 224
Analysis, vector, 145
Angles, Eulerian, 173
Angular momentum, 159
  conservation of, 26
Angular velocity, 164, 166
Arbitrary constant, 3
Areal velocity, 26
Argument of a complex number, 52
Auxiliary conditions in the calculus of variations, 96
Auxiliary equation, 40
Axes, principal, 170

B

Bessel's equation, 63
Betatron, 312
Body, rigid, 157
Brachistochrone, 88

C

Calorie, 188
Canonical distribution, 218
Canonical ensemble, 220
Canonical transformations, 113
Capacitance, 265
  to ground, 265
  mutual, 265
Cartesian coordinates, right-hand system of, 18
Center of mass, 26, 157
  separation of, 32
Characteristic equation of a thermodynamic system, 187
Charge density, 251
Chasles' theorem, 164
Coefficient, of mutual inductance, 320
  of self-inductance, 313, 320
Complementary function, 39
Complete set of orthogonal functions, 136
Complex conjugate, 51
Complex functions, 53
Complex numbers, 50
  representation of, 52
Components, orthogonal, 147
Conductance, 284
Conductivity, 284
Conductors, 263
Conic section, 35
Conservation, of density in phase, 217
  of energy, 23
Conservation laws of mechanics, 27
Constant, arbitrary, 3
Continuous string, forced vibrations of, 134
  normal vibrations of, 129
  traveling waves in, 142
Contraction, Lorentz, 350
Conversion table for electrical units, 326
Coordinates, Cartesian, 18
  normal, 83, 120
  orthogonal curvilinear, 240
Cosecant, hyperbolic, 54
Cotangent, hyperbolic, 54
Coulomb's law, 246
Couple, 161
Coupled oscillators, 80
Critical point, 193
Critically damped motion, 71

Cross product, 149
Curl, 233
Current, 284
  four-dimensional, 345
Current density, 285, 293
Currents, steady, vector potential due to, 294
Curvilinear coordinates, orthogonal, 240
Cyclical processes, 191

D

Damped vibrations, 67
Decay, modulus of, 69
Decomposition, principle of, 55
Definite integrals, 9
Degree of a differential equation, 4
Density, current, 293
  in phase, 216
    conservation of, 217
Diamagnetism, 303
Dielectric constant, 266
Dielectric susceptibility, 266
Differential equation, degree of, 4
  formulation of, 5
  homogeneous linear, 37
  linear, 8, 37
  nature of, 2
  order of, 4
  ordinary, 2
  partial, 4
Differential equations, elementary, 1
  with constant coefficients, linear, 40
Dipole, electric, 255
Directrix, 35
Displacement current, 323
Distribution, canonical, 218
  modulus of, 220
Divergence, 232
  of a tensor, 240
Dot product, 148
Dynamics of rigid bodies, 157

E

Eccentricity, 35
Electric dipoles, 255
Electric displacement, 261
Electric polarization, 256
Electrical units, 250, 290, 326
Electromagnetic field, 306
Electromagnetic field energy, 328
Electromagnetic field momentum, 330
Electromagnetic induction, 310
Electromagnetic potentials, 332
Electromagnetic waves, 334
Electromotive force, 287
Electrostatic field, 247, 306
Electrostatic system, energy of, 278
Ellipsoid of inertia, 171
Energy, absorption of, by an oscillator, 78
  associated with polarization, 279
  conservation of, 23
  of an electrostatic system, 278
  free, 209
  integral, 20
  kinetic, 21
  in a magnetic field, 303, 318
  and mass, equivalence of, 353
  potential, 21
  of a vibrating string, 133
Ensemble, canonical, 220
  microcanonical, 220
Enthalpy, 209
Entropy, 204
Equation, auxiliary, 40
  Bessel's, 63
  characteristic, 187
  Euler's, 172
  Euler-Lagrange, 93
  exponents of, 65
  homogeneous differential, 40
  indicial, 65
  Lagrange's, 106
  Legendre's, 61
  Maxwell's, 324
  nature of a differential, 2
  nonhomogeneous, 37
  ordinary differential, 2
  Poisson's, 269
  of state, 187
    van der Waals, 192
Equations, linear differential, 8
    with constant coefficients, 40
    with variable coefficients, 58

Equations, of motion, Newton's, 18
of order higher than the first, linear, 37
of the second order that do not contain y, 14
Equilibrium, statistical, 218
Equipartition theorem, 224
Equivalence of mass and energy, 353
Equivalent sets of forces, 162
Ergodic hypothesis, 221
Essential singularities, 63
Euler's equations, 172
Euler-Lagrange equation, 93, 106
Eulerian angles, 173
Exponents of a linear differential equation, 65
Extreme values, 89

F

Ferromagnetism, 303
Field, electromagnetic, 306
electrostatic, 247, 306
magnetostatic, 309
Field energy, electromagnetic, 328
Field momentum, electromagnetic, 330
Field strength, electric, 247
magnetic, 298
Fields, magnetic, due to currents only, 290
due to magnetization, 296
in moving coordinate systems, 313
First law of thermodynamics, 193
Fixed axis, rotation about, 177
Force, inverse-square, 32
moment of, 26, 160
Forces, equivalent sets of, 162
harmonic, 73
nonconservative, 110
sinusoidal, 73
Formula, recursion, 60
Formulation of the differential equation, 5
Four-dimensional current, 345
gradient, 345
potential, 345
Fourier series, 133, 137
Free energy, 209
Free rotation of a rigid body, 181
Function, complementary, 39
Gibbs', 209
Green's, 48
Hamiltonian, 113
Lagrangian, 105
partition, 227
Function notation, 4
Functions, complex, 53
hyperbolic, 54
orthogonal, 136

G

Galilean transformations, 339
Gauss's law, complete form of, 260
differential form of, 261
restricted form of, 253
Gauss's theorem, 236
Gaussian system of units, 250, 325
Gibbs' function, 209
Giorgi system of units, 250
Gradient, 230
four-dimensional, 345
of a tensor, 239
Gravitation, law of, 32
Green's function, 48

H

Hamilton's principle, 102
derivation of, 102
Hamilton-Jacobi partial differential equation, 113
Hamiltonian function, 113
Harmonic forces, 73
Harmonic motion, plane, 31
simple, 30
Harmonics, spherical, 273
Heat, latent, 189
mechanical equivalent of, 194
mechanical theory of, 185
quantity of, 188
specific, 189
Heat capacity, 189
Heat content, 209
Heat function, 209
Hermite polynomials, 137
Homogeneous linear differential equation, 37
solution of, 40

Hyperbolic cosecant, 54
Hyperbolic cotangent, 54
Hyperbolic functions, 54
Hyperbolic secant, 54
Hypothesis, ergodic, 221

I

Imaginary numbers, 51
Imaginary part, 51
Index of probability, 220
Indicial equation, 65
Induction, electromagnetic, 310
Inertia, ellipsoid of, 171
  moments of, 170
  products of, 170
  tensor of, 169
Insulation, vibration, 78
Insulators, 263
Integral, energy, 20
  line, 234
  particular, 39, 42
  surface, 235
Integrals, indefinite, 9
Integration, numerical, 10
Inverse-square force, 32
Irrational numbers, 51
Irregular points, 63
Isoperimetric problems, 98

J

Joule-Thomson porous-plug experiment, 211

K

Kinematics of a rigid body, 163
Kinetic energy, 21
  of a rotating body, 175

L

Lagrange's equations, 106
Lagrangian function, 105
Latent heat, 189
Law, Coulomb's, 246
  of motion, Newton's third, 24
  Ohm's, 284
Laws, Ampère's, 289
Legendre's equation, 61
Line integral, 234
Linear differential equations, with constant coefficients, 40
  general properties of, 37
Linear equations of order higher than the first, 37
Linear momentum, 157, 159
Linear vector functions, 154
Liouville's theorem, 217
Loaded string, vibrations of, 121
Logarithmic decrement, 69
Lorentz contraction, 350
Lorentz's form of Maxwell's equations, 335
Lorentz transformation, 342

M

Magnetic field, energy in, 318
Magnetic field strength, 298
Magnetic fields due to magnetization, 296
Magnetic moment, 293
Magnetic vector potential, 291
Magnetization, 283, 297
Magnetostatic field, 309
Magnetostatics, 283
Mass, center of, 26, 157
  and energy, equivalence of, 353
  transformation of, 351
Maxwell's equations, 324
  Lorentz's form of, 335
Maxwell stress tensor, 332
Mechanical equivalent of heat, 194
Mechanical theory of heat, 185
Mechanics, statistical, 213
Meter-kilogram-second-coulomb system of units, 250, 325
Methods of thermodynamics, 186
Microcanonical ensemble, 220
Modulus, of a complex number, 52
  of decay, 69
  of the distribution, 220
Moment, dipole, 255
  of force, 160
Moments of inertia, 170

Momentum, 25
angular, 26, 159
conservation of, 25
linear, 157, 159
moment of, 26
Momentum integral, 24
Motion, critically damped, 71
of the planets, 32
simple-harmonic, 30
Mutual capacitance, 265
Mutual inductance, coefficient of, 320

N

Newton's equations of motion, 18
Newton's third law of motion, 24
Nonconservative forces, 110
Nonhomogeneous differential equation, 37
Nonuniform string, 138
Normal coordinates, 83, 120
Normal vibration, 83
variation problem for, 141
Normalized functions, 136
Numbers, complex, 50
imaginary, 51
irrational, 51
real, 51
Numerical integration, 10

O

Ohm's law, 284
Operator D, 39
Order of a differential equation, 4
Ordinary point of a differential equation, 58
Orthogonal components of a vector, 147
Orthogonal coordinates, 241
Orthogonal curvilinear coordinates, 240
Orthogonal functions, 136
Orthogonal set, complete, 136
Oscillators, coupled, 80

P

Paramagnetism, 302
Partial differential equation, 4
Particular integral, 39
Partition function, 227
Pendulum with arbitrary amplitude, 113
Perfect gas, 191
Period, 69
Permeability magnetic, 303
Permittivity, 266
Phase, density in, 216
conservation of, 217
Phase integral, 227
Phase lag, 74
Phase space, 213
Plane simple-harmonic motion, 31
Planets, motions of the, 32
Poisson's equation, 269
Polarization, electric, 256
energy associated with, 279
Polynomials, Hermite, 137
Porous-plug experiment, Joule-Thomson, 211
Positive definite expressions, 120
Potential, 244, 247
electromagnetic, 332
four-dimensional, 345
magnetic vector, 291
scalar, 332
vector, 332
Potential energy, 21
Postulates of relativity, 340
Poynting's vector, 330
Practical system of units, 325
Principal axes, 170
Principle, Hamilton's, 102
variation, 87
Probability, index of, 220
Problems of thermodynamics, 186
Processes, cyclical, 191
Product, cross, 149
dot, 148
of inertia, 170
scalar, 148
vector, 149
Projectile, motion of a, 28

Q

Q of an oscillator, 70
Quantity of heat, 188

R

Real numbers, 51
Real part, 51
Recursion formula, 60
Regular points of a differential equation, 63
Relativity, postulates of, 340
  restricted theory of, 339
  of simultaneity, 349
Resistance, 284
Resonance, 75
Restricted theory of relativity, 339
Reversible processes, 190
Rigid bodies, dynamics of, 157
  free rotation of, 181
  kinematics of, 163
  laws of motion of, 157
Rotating body, kinetic energy of, 175
Rotation about a fixed axis, 177

S

Scalar potential, 332
Scalar product, 148
Scalar quantities, 145
Secant, hyperbolic, 54
Second law of thermodynamics, 201
Self-inductance, coefficient of, 320
Separation of variables, 5
Series, Fourier, 133
  of orthogonal functions, 136
Simple-harmonic motion, 30
Simpson's rule, 10
Singular points of a differential equation, 59
Singularities, of a differential equation, 59
  essential, 63
  regular, 63
Sinusoidal forces, 73
Solution, general, 3
  around ordinary points, 59
Specific heat, 189
Spherical harmonics, 273
State of a thermodynamic system, 186
Stationary value, 90
Statistical equilibrium, 218
  mechanics, 213
Steady currents, 294
Stokes's theorem, 238
Stress tensor, Maxwell, 332
String, continuous, 129
  loaded, 121
  nonuniform, 138
Sturm-Liouville equation, 138
Summation convention, 154
Superposition, principle of, 55
Surface integral, 235
Systems, vibrating, 120

T

Temperature, 188
  absolute scale of, 204
Tensor, 155
  divergence of a, 240
  field, 239
  gradient of a, 239
  of inertia, 169
Thermodynamic analogies, 224
Thermodynamic potential at constant pressure, 209
Thermodynamic system, 186
Thermodynamics, first law of, 193
  methods of, 186
  problems of, 186
  second law of, 201
  third law of, 186
Theorem, Chasles', 164
  equipartition, 224
  Gauss's, 236
  Liouville's, 217
  Stokes's, 238
  of uniqueness, 271
Theory of relativity, restricted, 339
Thermometry, 188
Third law, of motion, Newton's, 24
  of thermodynamics, 186
Torque, 160
Transformation, of mass, 351
  of velocity, 351
Transformations, canonical, 133
  Galilean, 339
Transient, 76
Traveling waves in a string, 142

U

Underdamped vibrations, 68
Uniqueness, theorem of, 271
Units, conversion table for, 326
  electrical, 250
  Gaussian, 250
  Gaussian system of, 325
  Giorgi system of, 250
  mks system of, 250

V

Value, of a complex number, absolute, 52
  extreme, 89
  stationary, 90
Van der Waals equation of state, 192
Variation, first, 93
Variation principle, 87
Variation problem for normal vibrations, 141
Vector, definition of, 145
  differentiation of, 151
  Poynting's, 330
Vector analysis, 145
Vector field, 230
Vector functions, linear, 154
Vector potential, 332
  magnetic, 291
  properties of, 294
Vector product, 149
Velocity, angular, 164, 166
  transformation of, 351
Vibrating string, energy of, 133
Vibrating systems, 120
Vibration, normal, 83
Vibration insulation, 78
Vibrations, damped, 67

W

Waves, electromagnetic, 334
  in a string, traveling, 142
Work function, 209

PRINCIPLES OF MATHEMATICAL PHYSICS

HOUSTON

SECOND EDITION

53.01
H81

McGRAW-HILL BOOK COMPANY